Genetic dissection of growth and wood properties in a nested, half-sib *Eucalyptus* hybrid pedigree

by

Julia Candotti

Submitted in partial fulfilment of the requirements for the degree

Magister Scientiae

In the Faculty of Natural and Agricultural Sciences

Department of Biochemistry, Genetics and Microbiology

University of Pretoria

December 2019

Under the supervision of Professor Alexander A. Myburg

and co-supervision of Professor Eshchar Mizrachi

Declaration

I, Julia Candotti declare that the dissertation, which I hereby submit for the degree MSc Genetics at the University of Pretoria, is my own work and has not previously been submitted by me for a degree at this or any other tertiary institution

…….

Julia Candotti

December 2019

Dissertation Summary

Genetic dissection of growth and wood properties in a nested, half-sib *Eucalyptus* hybrid pedigree

Julia Candotti

Supervised by **Prof. A.A. Myburg**

Co-supervised by **Prof E. Mizrachi**

Submitted in partial fulfilment of the requirements for the degree ***Magister Scientiae***

Department of Biochemistry, Genetics and Microbiology

University of Pretoria

Eucalyptus is important for the forestry industry due to its excellent growth and wood properties. In crop species, nested multi-parent populations have been used to increase the power and resolution of quantitative trait loci (QTL) detection. These populations have predominantly been used in species in which recombinant inbred lines can be generated and have not been fully exploited in outcrossing species such as *Eucalyptus*. To determine if multi-parent mapping approach can be used effectively for genetic dissection in *Eucalyptus*, we made use of an existing F_1 hybrid trial series, consisting nine *E. grandis* pollen parents and eight *E. urophylla* seed parents. The population has many full-sib (FS) families nested within half-sib (HS) families and was planted across four different sites.

The objectives of this MSc study were to i) construct genetic linkage maps of one *E. grandis* pollen parent and one *E. urophylla* seed parent of the multi-parent population, ii) analyse transmission ratio

distortion of mapped markers in the F_1 hybrid progeny to identify hybrid compatibility barriers, iii) map QTLs underlying growth and wood properties in the two pure species parental maps.

We constructed framework genetic linkage maps for the *E. grandis* pollen parent and the *E. urophylla* seed parent. A total of 388 (*E. grandis* HS family, n = 349) and 422 (*E. urophylla* HS family, n = 367) single nucleotide polymorphisms (SNP) markers were included in the linkage maps resulting in an average marker density of 2.4 cM. Using the genetic linkage maps, we identified 15 and 23 QTLs underlying growth and wood properties for the *E. grandis* and *E. urophylla* HS family, respectively. We identified large to medium effect QTLs, with the percentage of variance explained ranging from 3.06% to 36.58%. We identified different QTLs across the sites which suggests that the traits are affected by genotype-by-environment interaction. We analysed segregation distortion of the markers included in the framework genetic linkage maps within HS families, FS families and sites. We found that there is a large amount of segregation distortion (between 0 – 29.38% distortion) and that the patterns of distortion varied for individual FS families planted across multiple sites and single sites with multiple FS families. We were also able to identify potential pre- and postzygotic barriers to hybrid compatibility through the analysis of segregation distortion of dead and living trees. Taken together, these results show that there are both parent specific interactions, that are dependent on the environment, which underlie hybrid compatibility.

In this study, we applied an approach whereby genetic linkage maps can be constructed and QTL identified in an outcrossing multi-parent mapping population. We show that multi-parent populations hold promise for studying hybrid compatibility, as diverse founders are crossed resulting in a number of F_1 hybrid progeny. The results of this study show that this approach can be applied in existing F_1 hybrid breeding trials for more fine scale genetic dissection of complex trait variation as well as hybrid compatibility of *E. grandis* and *E. urophylla*.

Preface

Eucalyptus is an important tree genus for hardwood plantation forestry due its desirable growth and wood properties. It is commonly planted as interspecific hybrid clones to combine the favourable characteristics of two species into a single genetic background. *E. grandis* has desirable wood and growth properties, but is highly susceptible to fungal diseases. Therefore, *E. grandis* is commonly crossed with *E. urophylla,* a tropical eucalypt species, that is more disease resistant. Despite the success of the interspecific hybrid breeding approach, many specific parental crosses do not yield any hybrid progeny. This can be due to pre- and postzygotic incompatibilities, the genetic factors of which can be fixed or segregating in the parental species. Identification of loci underlying these incompatibilities are important for breeding programmes as it will allow for breeders to determine which trees to cross.

Interspecific hybrids are often used for genetic map construction because they maximise genetic diversity available for linkage analysis. The first genetic linkage maps were constructed for *E. grandis* and *E. urophylla* interspecific hybrids in 1994 (Grattapaglia *et al.* 1994). Since this study, many genetic linkage maps have been constructed. However, due to the high linkage disequilibrium in biparental crosses and a low marker density, the genetic linkage maps had low resolution for quantitative trail loci (QTL) detection. In 2013, the first genome-wide association study (GWAS) was reported in *Eucalyptus* (Cappa *et al.* 2013). However, due to the experimental design and small population size, this study had a low resolution and power to detect marker-trait associations. Due to a limited number of genomic resources available for *Eucalyptus*, there have been few subsequent GWAS studies as a high marker coverage could not be achieved for a full GWAS in a population with low linkage disequilibrium (LD).

Many studies in crop species have also experienced the limitations with linkage analysis and GWAS. Therefore, in 2008, the first multi-parent mapping populations were designed in maize to combine the high power of linkage analysis in single families with the high resolution of GWAS at population level (Yu *et al.* 2008). These populations are also advantageous as they result in a balanced representation of allele variation as all alleles segregate in at least one family. This results in an enhanced power and precision for estimating allelic effects. Since the first study, a number of different types of multi-parent population designs have emerged. Despite differences in the designs, the basic principle is the same, which is to cross a number of diverse founders and generate recombinant inbred lines (RILs). Studies in multi-parent populations have shown to have high power and high resolution to detect QTL. However, these studies have been limited to species in which RILs can be generated and have not been fully explored in outcrossing genera such as *Eucalyptus,* despite the fact that multi-parent crossing schemes are commonly used in *Eucalyptus* breeding, especially hybrid breeding trials.

Recent advances in the development of *Eucalyptus* resources has opened the doors for genomic dissection studies. The completion of the reference genome sequence of *Eucalyptus grandis* in 2014 (Myburg *et al.* 2014) enabled the generation of the *Eucalyptus* EUChip60K SNP chip (Silva-Junior *et al.* 2015). This has provided a platform for large-scale genotyping of *Eucalyptus* populations. Additionally, the crossing of multiple parents is a common practice in F_1 hybrid breeding trials for *Eucalyptus* but these trials have not been fully explored for genetic dissection of complex traits. Sappi Forest Research (Hilton, KZN, South Africa) have provided access to such a breeding trial that can be used for half-sib (HS) family mapping with nested F_1 full-sib (FS) families, replicated over sites. The trial was generated by crossing nine *E. grandis* pollen parents with eight *E. urophylla* seed parents. This population therefore has a number of full-sib families nested within half-sib families. This provides the opportunity to dissect quantitative traits across full and half-sibs. The population

design also allows for the analysis of hybrid incompatibility as a large number of diverse founders were crossed and not all the crosses yielded viable progeny.

The overall aim of this MSc was to determine the genetic architecture of growth and wood properties segregating from pure species parents in F_1 hybrid progeny of *E. grandis* and *E. urophylla* and to determine the genome-wide architecture of hybrid incompatibility between parental genomes. To evaluate the possibility to achieve this in the context of a multi-parent F_1 hybrid breeding trial, we had the following objectives: (1) construct genetic linkage maps for one *E. grandis* pollen parent and one *E. urophylla* seed parent (2) analyse segregation distortion of SNP markers in the F_1 hybrid progeny and (3) map QTL controlling growth and wood properties in the two pure species parental maps. We were able to apply existing methods to construct genetic linkage maps and identify QTL in an outcrossing multi-parent mapping population. Furthermore, we were able to use segregation distortion analysis to identify regions of the parental genomes potentially underlying hybrid incompatibility.

Chapter 1 of this dissertation provides an overview of quantitative and hybrid genetics with a focus on plants. For the quantitative genetics sections I discuss the advantages and disadvantages of previous methods that have been used to dissect quantitative traits, and then how recent advances in population design has allowed for the advantages of previous methods to be combined in a single method. In addition, I review the theory behind heterosis and hybrid incompatibility to determine where we are with understanding the genetics underlying these mechanisms. I then give an overview of *Eucalyptus* hybrid breeding and genetic dissection studies in *Eucalyptus*. Furthermore, I discuss recent advances in the development of *Eucalyptus* genetic and genomic resources with focus on how these can be used to improve our understanding of *Eucalyptus* quantitative traits and hybrid compatibility.

In **Chapter 2**, I describe the construction of genetic linkage maps in a *Eucalyptus* F_1 multi-parent mapping population. Furthermore, I identify quantitative trait loci (QTL) underlying growth and wood properties across different environments and show that genotype-by-environment interactions affect the identified QTLs. The results demonstrate that multi-parent mapping approaches can be used in outcrossing plants such as *Eucalyptus* to identify marker-trait associations.

For **Chapter 3**, I analyse the segregation distortion patterns of markers included in the parental genetic linkage maps from Chapter 2. Due to the population design, segregation distortion analysis is performed within half-sib families, full-sib families and sites. I show that this can be used to identify specific interactions between the parental genomes which are also dependent on interactions with the environment. Furthermore, I analyse the segregation patterns of dead and living trees within an intersecting full-sib family (sharing the pollen and seed parent for which genetic linkage maps were constructed) which enables us to identify of regions of the parental genome underlying potential pre- and postzygotic incompatibilities. Overall, the results of this chapter show that there are complex genetics underlying hybrid incompatibility which multi-parent mapping approaches can start to identify and resolve.

This MSc dissertation was undertaken from January 2018 to December 2019 in the Department of Biochemistry, Genetics and Microbiology and the Forestry and Agricultural Institute (FABI) at the University of Pretoria. This study was completed under the supervision of Prof. A.A. Myburg and co-supervised by Prof. E. Mizrachi. The multi-parent mapping population used in this study was constructed and maintained by Sappi Forest Research (Hilton, KZN, South Africa). Chapter 2 of this study has been prepared in the format of an independent manuscript for submission to a peer-reviewed journal (*Tree Genetics and Genomes*).

Preliminary results of this MSc have been presented at the following national and international conferences:

International:

Julia Candotti, Marja M. O'Neill, S. Melissa Reynolds, Roobavathie Naidoo, Nicoletta Jones, Eshchar Mizrachi, Alexander A. Myburg, Genetic dissection in a multi-parent *Eucalyptus* F1 hybrid mapping population, 23 - 28 June 2019, Raleigh,NC, USA (Poster).

Candotti J, O'Neill MM, Reynolds SM, Naidoo R, Kanzler A, Jones N, Mizrachi E. Myburg AA. 2019. Genetic dissection of growth and wood development in nested, multi-parent *Eucalyptus* hybrid populations. Eucalypt genetics: fundamental and applied research in a post-genome era, 18-21 February 2019, Hobart, Tasmania, Australia (Speed Talk).

National:

Candotti, J., O'Neill, M.M., Reynolds, S.M., Naidoo, S., Kanzler, A., Jones, N., Mizrachi, E., and Myburg, A.A. 2018. Towards nested, multi-parent genetic dissection of growth and wood development in *Eucalyptus* hybrid populations. South African Society for Bioinformatics and South African Genetics Society conference, "Sequence anlaysis from this generation to the next", 16-18 October 2018, Golden Gate resort, Free State, South Africa (Poster).

Acknowledgements

I would like to express my sincere appreciation and gratitude to the following people and institutes for their involvement and support throughout this MSc:

- Prof Zander Myburg for your supervision, guidance and encouragement throughout this study. Thank you for giving me the opportunity to do this study and for always being willing to explain concepts to me. The conferences and workshops which you arranged for me to attend have proved invaluable in improving my understanding and increasing my subject knowledge and your support has been much appreciated. Additionally, thank you for making FMG a great place to carry out research and for encouraging group interaction through socials and field tips.
- Prof Eshchar Mizrachi for your contribution to this study and for always asking me those tricky hypothesis questions in my committee meetings. You encouraged me think about my project from different perspectives.
- Mrs Marja O'Neill and Ms Melissa Reynolds (Marjalissa) for supporting me throughout this project and always being willing to help me. Thank you for always putting up with my silly questions and for having the patience to explain concepts multiple times over. Your assistance with my writing and presentations is greatly appreciated. You have provided me with excellent advice and I have learnt so much from you regarding how best to present my research. Finally, thank you for all the jokes and laughs which always brighten my day.
- Ms Lizette Loubser for your friendship, good humour and all your help throughout this project. Your help on the bioinformatics side of the project was invaluable, I would not have been able to do it without you! Thank you for reading my writing and providing constructive feedback. Finally, thank you also for always being hilarious and making me laugh.

- Mr Luke Kim for your constant encouragement, support and friendship. Thank you for being the best office buddy ever and always making me laugh! Thank you also for giving me helpful feedback on my writing. I am extremely grateful for your support throughout this study.
- Ms Thandeka Ngondo for always putting a smile on my face and for your friendship. Thank you for being an amazing friend who is always willing to help no matter what.
- Members of the Forest Molecular Genetic Programme for the support. From friendly smiles in the corridor to valuable feedback from presentations and peer review, you have provided a culture of encouragement and support. I would especially like to thank the following people for their contributions and support:
    - Prof Sanushka Naidoo for being on my project committee and for your valuable input into the project.
    - Dr Nanette Christie for the workshops on R and statistics, the skills that these workshops have taught me are invaluable. Additionally, thank you for making the circos images for me.
    - Mrs Patience Motaung for teaching me DNA extractions.
    - Ms Tersia Maobelo for teaching me how to perform parentage analysis.
- Sappi Forest Research for providing the plant material for this study and for funding the project
- Department of Science and Technology (DST), Technology Innovation Agency (TIA) and Technology and Human Resources for Industry Programme (THRIP) for funding the project.
- National Research Foundation (NRF) for the bursary and funding the project.
- Department of Biochemistry, Genetics and Microbiology (BGM) and the Forestry and Agricultural Biotechnology Institute (FABI) at the University of Pretoria for providing the facilities and an excellent research environment.
- My parents and sister for their unlimited love encouragement and support. I am extremely grateful and appreciate everything you have done for me.

Table of Contents

Declaration ii

Dissertation Summary iii

Preface v

Acknowledgements x

Table of Contents xii

List of Tables xvi

List of Figures xvii

List of Supplementary Tables xviii

List of Supplementary Figures xx

List of Supplementary Files xxii

Chapter 1: Literature Review: Methods for genetic dissection of quantitative traits and hybrid compatiblity in plants 1

1.1 Introduction 2

1.2 Molecular genetic dissection of quantitative traits 4

1.2.1 Linkage analysis 4

1.2.2 Genome-wide association studies 6

1.2.3 Multi-parent mapping populations 7

1.3 Hybrid genetics 9

1.3.1 Heterosis 9

1.3.2 Hybrid incompatibility 10

1.4 *Eucalyptus* 12

1.4.1 *Eucalyptus* domestication and hybrid breeding 12

1.4.2 Genetic dissection studies in *Eucalyptus* 14

1.4.3 Genetic and genomic resources available for *Eucalyptus* 17

1.5 Conclusion 18

1.6 Figures 20

1.7 Tables 26

1.8 References 28

Chapter 2: Identification of QTL underlying growth and wood properties in a nested, multi-parent *Eucalyptus* hybrid population 33

2.1 Abstract 34

2.2 Introduction 34

2.3 Methods 37

2.3.1 Plant material and DNA isolation 37

2.3.2 Trait data 37

2.3.3 Parentage analysis 38

2.3.4 SNP genotyping 39

2.3.5 Species discrimination 39

2.3.6 Identification of informative SNPs 40

2.3.7 Genetic map construction 40

2.3.8 QTL mapping 41

2.4 Results 42

2.4.1 Parentage confirmation 42

2.4.2 Species identification 42

2.4.3 SNP genotyping and identification of informative SNPs 43

2.4.4 Genetic linkage maps 43

2.4.5 Trait data 45

2.4.6 QTL mapping 45

2.5 Discussion 47

2.5.1 Identification of informative markers 48

2.5.2 Genetic linkage maps 48

2.5.3 QTL mapping 50

2.5.4 Genotype-by-environment effect on QTL 52

2.6 Conclusions 53

2.7 Acknowledgements 54

2.8 Tables 55

2.9 Figures 62

2.10 References 66

2.11 Supplementary Figures 69

2.12 Supplementary Tables 105

Chapter 3: Analysis of hybrid incompatibliy in a *Eucalyptus* multi-parent population 119

3.1 Abstract 120

3.2 Introduction 120

3.3 Materials and Methods 125

3.3.1 Plant material and SNP genotyping 125

3.3.2 Segregation distortion analysis 125

3.4 Results 126

3.4.1 Segregation distortion within HS families 126

3.4.2 Segregation distortion within FS families 126

3.4.3 Segregation distortion within sites 127

3.4.4 Segregation distortion across site and FS family 127

3.4.5 Dissemination of pre- and postzygotic incompatibilities 128

3.5 Discussion 129

3.5.1 Identification of segregation distortion 131

3.5.2 Causes of segregation distortion 133

3.5.3 Application for industry 135

3.6 Conclusion and future prospects 136

3.7 Acknowledgments 137

3.8 Tables 138

3.9 Figures 139

3.10 References 146

3.11 Supplementary Figures 149

3.12 Supplementary Tables 162

Chapter 4 163

Concluding Remarks 163

List of Tables

Table 1.1 Summary of genetic linkage maps constructed for *E. grandis* 26

Table 1.2 Summary of genetic linkage maps constructed for *E. urophylla* ……..... 27

Table 2.1 *Eucalyptus* multi-parent mapping population.. 55

Table 2.2 Summary of SNP markers mapped for each linkage group in the *E. grandis* HS family 56

Table 2.3 Summary of SNP markers mapped for each linkage group in the *E. urophylla* HS family57

Table 2.4 Summary statistics of the corrected trait data of the *E. grandis* HS family. 58

Table 2.5 Summary statistics of the corrected trait data of the *E. urophylla* HS family. 59

Table 2.6 QTL detected for the *E. grandis* HS family pollen parent. QTL analysis was performed across three sites and the entire HS family.. 60

Table 2.7 QTL detected for the *E. urophylla* HS family seed parent. QTL analysis was performed across two sites and across the HS family.. 61

Table 3.1 Summary of significantly distorted markers for each FS family per site.. 138

List of Figures

Figure 1.1 Bi-parental cross used for linkage analysis. 20

Figure 1.2 Natural population used for Genome-Wide Association Studies (GWAS). 21

Figure 1.3 Multi-Advanced Generation Inter-Cross (MAGIC) population. 22

Figure 1.4 Nested Association Mapping (NAM) population. 24

Figure 1.5 Pre- and post-zygotic barriers to hybrid compatibility. 25

Figure 2.1 Framework genetic linkage map with QTL segregating in the *E. grandis* HS family pollen parent. 63

Figure 2.2 Framework genetic linkage map with QTL segregating in the *E. urophylla* HS family. 65

Figure 3.1 Summary of segregation distortion for each FS family within the *E. grandis* HS family. 140

Figure 3.2 Summary of segregation distortion for the *E. grandis* HS across four sites. 141

Figure 3.3 Segregation distortion magnitude, direction and distribution of a single site with multiple FS families and a single FS family across multiple sites. 143

Figure 3.4 Identification of regions underlying pre- and post-zygotic incompatibilities based on segregation distortion patterns of dead and alive trees. 145

List of Supplementary Tables

Supplementary Table 2.1 Site information for the four sites across which the multi-parent population was planted. 105

Supplementary Table 2.2 Colony run settings. 106

Supplementary Table 2.3 Markers which mapped to a different linkage group in the *E. grandis* genetic linkage map compared to *E. grandis* V2 genome assembly. 107

Supplementary Table 2.4 *E. urophylla* markers which mapped to a different linkage group compared to the chromosome position in the *E. grandis* V2 reference genome assembly. 108

Supplementary Table 2.5 Summary statistics of the raw trait data of the *E. grandis* HS family. 109

Supplementary Table 2.6 Summary statistics of the raw trait data of the *E. urophylla* HS family.. 110

Supplementary Table 2.7 Two-way ANOVA of phenotypic data for the *E. grandis* HS family. 111

Supplementary Table 2.8 One-way ANOVA of the phenotypic data, with FS family as the condition tested for the *E. grandis* HS family across sites.. 112

Supplementary Table 2.9 Two-way ANOVA of raw data for the *E. urophylla* HS family using family, site and family within site as the conditions. 113

Supplementary Table 2.10 One-way ANOVA for phenotypic data of the *E. urophylla* HS family across site. 114

Supplementary Table 2.11 Two-way ANOVA of corrected phenotypic data for the *E. grandis* HS family. 115

Supplementary Table 2.12 One-way ANOVA for corrected phenotypic data for the *E. grandis* HS family across site. 116

Supplementary Table 2.13 Two-way ANOVA of corrected phenotypic data for the *E. urophylla* HS family. 117

Supplementary Table 2.14 One-way ANOVA of corrected data for the *E. urophylla* HS family. 118

Supplementary Table 3.1 Number of individuals per FS family, of the *E. grandis* HS family, on each of the four sites. 162

Supplementary Table 3.2 Number of individuals per FS families, of the *E. urophylla* HS family, on each of the four sites. 162

List of Supplementary Figures

Supplementary Figure 2.1 PCA showing the clustering of SNP genotypes of FS families in the *E. grandis* and *E. urophylla* HS families.....69

Supplementary Figure 2.2 *E. grandis* full genetic linkage map and physical map.....73

Supplementary Figure 2.3 *E. urophylla* full genetic linkage map and physical map. 78

Supplementary Figure 2.4 Framework genetic linkage map and physical map for the *E. urophylla* seed parent. 79

Supplementary Figure 2.5 Framework genetic linkage map and physical position map for *E. grandis* pollen parent. 81

Supplementary Figure 2.6 Trait distribution for the *E. grandis* HS family across the different sites. 84

Supplementary Figure 2.7 Trait distribution for the *E. urophylla* HS family across the different sites. 86

Supplementary Figure 2.8 Phenotypic correlations for all corrected trait data. 88

Supplementary Figure 2.9 QTL profiles for *E. grandis* HS family jointly and across three sites..95

Supplementary Figure 2.10 QTL profiles for the *E. urophylla* HS family jointly and across two sites. 102

Supplementary Figure 2.11 Trait data and number of individuals of each genotypic class for growth QTL detected in regions of significant segregation distortion.....104

Supplementary Figure 3.1 Segregation distortion magnitude, direction and genome-wide distribution for the *E. urophylla* FS and HS families. 150

Supplementary Figure 3.2 Segregation distortion magnitude, direction and genome-wide distribution for the *E. urophylla* HS family across the four different sites.....152

Supplementary Figure 3.3 Segregation distortion magnitude, direction and genome-wide distribution for each FS family in the *E. grandis* HS family across four sites. 154

Supplementary Figure 3.4 Segregation distortion magnitude, direction and genome-wide distribution of each FS family in *E. grandis* HS family, across the different sites. 157

Supplementary Figure 3.5 Segregation distortion magnitude, direction and genome-wide distribution of each FS family in *E. urophylla* HS family, across the four sites 159

Supplementary Figure 3.6 Segregation distortion magnitude, direction and genome-wide distribution for each FS family in *E. urophylla* HS family, across four sites.....161

List of Supplementary Files

Supplementary File 2.1 Informative markers for the *E. grandis* and *E. urophylla* HS families.

Supplementary File 2.2 Markers with 100% similarity.

Supplementary File 2.3 Markers removed due to 100% similarity.

Supplementary File 2.4 Markers included in the full and framework genetic linkage maps for the *E. grandis* HS family.

Supplementary File 2.5 Markers included in the full and framework genetic linkage maps for the *E. urophylla* HS family.

Supplementary File 2.6 Input files for QTLCartographer

Supplementary File 3.1 Segregation distortion calculations for the *E. grandis* HS family and FS families across all sites.

Supplementary File 3.2 Segregation distortion calculations for the *E. urophylla* HS family and FS families across all sites.

Supplementary File 3.3 Segregation distortion calculations for the *E. grandis* HS family within each of the four sites.

Supplementary File 3.4 Segregation distortion calculations for the *E. urophylla* HS family within each of the four sites.

Supplementary File 3.5 Segregation distortion calculations for individual *E. grandis* FS families within each of the four sites.

Supplementary File 3.6 Segregation distortion calculations for individual *E. urophylla* FS families within each of the four sites.

Supplementary File 3.7 Segregation distortion calculation for dead and living trees in the intersecting FS family.

Chapter 1: Literature Review

Methods for genetic dissection of quantitative traits and hybrid compatibility in plants

1.1 Introduction

Genomics is an integral part of understanding the relationship between phenotype and genotype, which in turn is important for plant breeding programmes. With an increase in the human population, more resources are required from plants, such as food, wood and biofuels. To optimally obtain these products, efficient plant breeding programmes need to be in place. In order to do this, an understanding of the genetics underlying important traits is required. Many agronomically important traits are complex and controlled by multiple quantitative trait loci (QTL). QTLs have the ability to interact with each other as well as with the environment to cause variation in the phenotype (Mackay 2001). It is therefore important to understand both the genetics underlying a trait of interest as well as the influence of the environment on the trait. Genomics can be used to identify loci underlying a trait of interest.

Marker-trait associations are used to identify QTLs underlying complex traits and the information can be used for marker-assisted selection (MAS, Collard *et al.* 2005). MAS is advantageous as it allows for the tracking and combining of favourable alleles as well as the ability to monitor the genetic diversity present in the population. The application of MAS in breeding programs is challenging because it depends on the heritability of the trait, the genetic architecture of the trait and the number of loci affecting the trait (Abiola *et al.* 2003). The method by which marker-trait associations are detected also plays an important role in the success of MAS. Two commonly used methods for identifying marker-trait associations are linkage analysis and genome-wide association studies (GWAS). However, due to limitations of these methods, their application in breeding programmes have been limited. A more recent experimental design, multi-parent populations, have combined the advantages of both linkage analysis and GWAS and show promise for improvements in MAS.

Hybrid populations are commonly used to identify marker-trait associations and are used in breeding programmes. Interspecific hybrids are advantageous because they allow for the combining and

tracking of favourable alleles into single genetic backgrounds. Hybrids commonly exhibit heterosis, which is where they outperform their parents. This makes hybrids important for crop species as they have the potential to improve crop yields. However, not all hybrid combinations are successful due to hybrid incompatibility. Genetic dissection studies can be used to identify marker-trait associations with loci underlying hybrid incompatibility. The results can then be used to determine which parental genotypes to combine to yield the best hybrid progeny. Despite the importance of hybrids in breeding programmes, the underlying causes of heterosis and hybrid compatibility are not fully understood.

Eucalyptus is widely planted for commercial purposes due to its growth and wood properties. In commercial plantations, *Eucalyptus* is commonly planted as hybrids to allow for the combination of desirable characteristics from two species to be present in a single genetic background. Due to the high heterozygosity and limited genomic resources of *Eucalyptus*, genetic dissection studies have had a limited success in breeding programmes. Recent advances in genomic resources for *Eucalyptus*, such as the completion of the reference genome (Myburg *et al.* 2014), the development of the EUChip60K SNP chip (Silva-Junior *et al.* 2015) and the development of a multi-parent population, will allow for an improved understanding of complex traits and hybrid genetics of *Eucalyptus*.

This review will focus on approaches used to identify QTLs in plant species and the use of hybrids in breeding programmes. We will discuss the advantages and disadvantages of linkage analysis and GWAS and how the advantages can be combined in multi-parent populations. An overview of multi-parent populations as well as their limitations will be given We will briefly discuss hybrid (in)compatibility, but as there are many reviews on this topic we will not go into a great amount of detail. At the end, an overview of how these methods have been applied in *Eucalyptus* will be given. The focus on *Eucalyptus* will shed light on why genetic dissection studies in *Eucalyptus* are lagging behind other crop species and the efforts that are being made to advance the field.

1.2 Molecular genetic dissection of quantitative traits

1.2.1 Linkage analysis

The traditional approach to identify the location of QTLs is through linkage analysis. Genetic linkage maps are first constructed by analysing the recombination frequency between genetic markers (Pierce 2014), which is used to determine the genetic distance (measured in centimorgan, cM) between markers. The markers are then ordered in the genetic linkage maps based on the genetic distances between them. One recombination event in 100 meiosis will result in a genetic distance of 1 cM (1 cM = 1% recombination). Therefore, the higher the recombination frequency between two markers, the further away the markers will be from each other in genetic distance.

The genetic linkage maps are then used to identify QTLs underlying traits of interest relative to marker positions in the maps, in a process known as QTL mapping. The principle behind QTL mapping is discussed in detail by Collard *et al.* 2005. Briefly, if a marker is found to be associated with a trait, the QTL underlying the trait will be near the marker in the genome. This is because markers that are in linkage with the QTL will segregate together. Linkage analysis has the ability to identify many regions of the genome underlying a trait of interest because the whole genome is analysed at once. This makes it well suited to identify regions underlying complex traits, which are controlled by multiple loci.

Linkage analysis is typically performed using biparental crosses, where two diverse parents are crossed and the F_1 progeny are analysed (Figure 1.1). An advantage of these populations is the high power to detect marker-trait associations, due to only two parental alleles segregating in a 50:50 ratio within the population (Mackay 2001). Therefore, if allele is linked to a QTL, a strong association can be made as half the progeny will contain the allele and QTL. This increases the probability of detecting an allele which is rare in a natural population, as it will be present in 50% of the progeny in a biparental cross. Although the segregation of two alleles allows for a high power to detect marker-

trait associations, it limits the diversity analysed to that present in the parents (Flint-Garcia *et al.* 2003). This is not representative of the genetic diversity present within the species, which can result in many QTLs not being detected. This has also caused the results of studies to differ due to different parents being used.

Biparental populations have large linkage disequilibrium (LD) blocks due to the limited amount of recombination (Flint-Garcia *et al.* 2003). The large LD blocks are both advantageous and disadvantageous for genetic map construction. The first advantage is that a low marker density is required in order to capture all of the LD blocks. This is especially advantageous in non-model organisms in which the identification of markers can be challenging. However, this is largely a historical problem as new state-of-the-art technologies, such as single nucleotide polymorphisms (SNP) chips and third-generation sequencing, allow for maker identification in non-model organisms. The second advantage is a high power to identify marker-trait associations. Due to the large LD blocks, a marker on one side of an LD block will be linked to a QTL on the opposite side of the LD block which will allow for an association between the marker and QTL, even though they may be far apart (Collard *et al.* 2005). However, the disadvantage of large LD blocks, is a low resolution because the genomic regions where there is an association with a trait are large (Mackay 2001). Therefore, a QTL can be identified in a large region, but the exact location, or causative gene, cannot be determined.

One way in which the resolution of biparental populations can be increased is through the use of F_2, backcross and Advanced Intercross Lines (AIL). In F_2 and backcross lines, recombination is increased through additional rounds of meiosis. This results in progeny with smaller LD blocks. In AIL, the F_1 are intercrossed for a few generations to increase the amount of recombination (Darvasi and Soller 1995). While, AIL does increase the amount of recombination, the population size needs to be at least 100 individuals and is limited to species which have short generation times as many

generations are required for the intercrossing step. While F_2, backcross and AIL can increase the amount of recombination to some extent, these populations are still derived from two parents which has its own limitations.

1.2.2 Genome-wide association studies

Association mapping is another approach used to identify markers associated with traits. Candidate gene association was first used to fine map QTLs following linkage analysis (Risch and Merikangas 1996). Here, a QTL region first needs to be identified using linkage analysis and then the QTL region is analysed separately with a larger number of markers. The theory behind association mapping is similar to linkage analysis in that a marker which is close to a QTL will be more likely to segregate with the QTL during recombination. Therefore, by having a higher marker density, the position of the QTL can be determined more accurately. Candidate gene association was initially used as large numbers of markers could not be identified genome-wide. With advances in technology, large numbers of polymorphic markers could be identified genome-wide and this led to the development of genome-wide association studies (GWAS).

In GWAS, markers across the entire genome are analysed to determine if they are associated with a trait (Hirschhorn and Daly 2005). An advantage of GWAS, is that it is performed in large, natural populations which have a high amount of genetic diversity and historical recombination (Figure 1.2). The high genetic diversity is advantageous as it allows for a large amount of variation to be analysed, which is more representative of the species (Zhu *et al.* 2008). Due to the large amount of historical recombination, the LD blocks are small, which results in a higher resolution for identifying QTLs. However, GWAS does not come without limitations of its own. The first limitation is the low power to detect markers associated with a trait (Mackay *et al.* 2009). This is due to alleles having a low frequency in natural populations with large amounts of genetic diversity. This is especially true for alleles which are rare in a population. In order to increase the power, larger samples sizes are required

which are often not feasible to obtain and genotype (Hirschhorn and Daly 2005). Population substructure of natural populations, is another limitation in GWAS. Population substructure can lead to an allele being at a higher frequency in a population subgroup. This can result in a false-positive association between the allele and a trait (Lander and Schork 1994; Hirschhorn and Daly 2005). Therefore, the effects of population substructure need to be detected and corrected prior to association analysis.

1.2.3 Multi-parent mapping populations

In the past ten years, multi-parent mapping populations have been constructed to combine the high power of linkage analysis and high resolution of GWAS. These populations have many advantages besides the high power and resolution, such as the control of population structure and the ability to analyse a large amount of variation through the use of diverse founders (McMullen *et al.* 2009). They are also advantageous as recombinant inbred lines (RILs) are generated which results in a large amount of recombination and also an eternal resource for studying genetic dissection. While these populations have been successful in many plants, they have been limited to species in which self-fertilization can take place. Multiparent Advanced Generation Intercross (MAGIC) and Nested Association Mapping (NAM) are two common designs of multi-parent populations.

MAGIC populations are constructed by intercrossing a large number of diverse founders, followed by intercrossing of the progeny for a few generations and then generating RILs (Figure 1.3). MAGIC lines of *Arabidopsis thaliana* were one of the first multiparent mapping populations constructed in plants (Kover *et al.* 2009). The population was constructed through four generations of intermating between 19 accessions, followed by six generations of inbreeding. The population design limits the amount of population structure and any population structure present, had a limited effect on QTLs. MAGIC lines are ideal for QTL analysis as they show a large amount of phenotypic and genetic variation as well as high recombination rates. Kover *et al* (2009), were able to use the MAGIC lines

to map QTLs with a higher accuracy and resolution when compared with previous biparental and Recombinant Inbred Lines (RIL) populations. Since the first MAGIC population, many variations of these mapping populations have been constructed in wheat (Huang *et al.* 2012), rice (Bandillo *et al.* 2013), tomato (Pascual *et al.* 2015), barley (Sannemann *et al.* 2015) and maize (Dell'Acqua *et al.* 2015).

Nested Association Mapping (NAM) is another type of multiparent mapping population (Figure 1.4). The first NAM population was constructed in maize by crossing 25 diverse founders to a single founder and the progeny were inbred for 5 generations (Yu *et al.* 2008). The maize NAM population, which consisted of 5000 RIL, had a high level of allelic richness and a large amount of recombination. The LD blocks were small due to the generation of RILs and the historical recombination present in the diverse founders. The maize NAM population had some population substructure present, but it was found that when the structure was ignored, the risk of false-positives did not increase due to balanced families being produced and the shuffling of founder genomes during RIL development. NAM was first used to determine the genetic architecture underlying flowering time in maize (Buckler *et al.* 2009). It was found that joint linkage QTL analysis across all the families, was able to identify almost twice as many significant effects when compared to the single-family (biparental) analyses. NAM has since been used to successfully identify QTLs for many traits in maize. Since the initial maize NAM study, this population type has been used in barley (Maurer *et al.* 2015), wheat (Bajgain *et al.* 2016), rice (Fragoso *et al.* 2017), sorghum (Bouchet *et al.* 2017) and soybean (Song *et al.* 2017). These NAM studies have been successful in identifying QTLs controlling many traits and show that NAM can be successfully applied to different types of crops.

1.3 Hybrid genetics

1.3.1 Heterosis

Heterosis occurs when hybrid progeny display improved traits when compared to the parents. This was first observed in 1908 by Shell (1908), who found that inbred maize lines had reduced vigour and growth properties. When these inbred lines were intercrossed, the vigour and growth properties of the hybrids exceeded that of the parents. Through the years, many crosses between different parents were performed to try to better understand how heterosis occurs. There are three hypotheses for the genetic basis of heterosis namely dominance, overdominance and epistasis. Dominance occurs when the superior allele mask the effect of the alternative allele in heterozygotes, while overdominance occurs when the heterozygote progeny is superior to the homozygote parents (Pierce 2014). Epistasis is the interaction between different loci and results in heterosis when the alleles interact in a favourable manner. Despite the advances in technology, the mechanism behind heterosis is still not fully understood.

Different combinations of the three mechanisms have been identified to play a role in heterosis. In a study using an immortalized F_2 population of maize, 13 heterotic loci for grain yield and its various components were identified (Tang *et al.* 2010). They were also able to identify 143 digenic interactions which showed that both dominance and epistasis played a role in heterosis for grain yield and its components. In another study, the genetic basis of rice yield was analysed in an immortalized F_2 population (Zhou *et al.* 2012). They analysed both epistasis and single-locus effects contributing to heterosis genome-wide. The results showed that the underlying cause of heterosis was trait-specific with overdominance affecting heterosis of yield while epistasis affected tillers per plant. Both of these mechanisms were found to be important in grain weight heterosis. These studies show that all three mechanisms can underlie heterosis which suggests that heterosis is caused by many complex interactions and the mechanism may be trait and species-specific.

1.3.2 Hybrid incompatibility

Despite hybrids commonly exhibiting heterosis, some hybrid combinations are incompatible. This can result in no hybrid progeny and hybrid necrosis. In this review we will focus on hybrid incompatibility and its effect on hybrid breeding programmes, however, speciation is an important phenomenon which readers should take into account and readers are directed to a review of plant speciation by Rieseberg and Blackman (2010). Hybrid incompatibility can either be prezygotic or postzygotic (Figure 1.5). Prezygotic mechanisms include flowering time and colour (Rieseberg and Blackman 2010), habitat, temporal barriers and pollen tube formation and growth rate (Snow *et al.* 2000; Rieseberg and Blackman 2010). Prezygotic barriers in plants are challenging to study as allele frequencies are required before and directly after fertilisation which had limited the studies performed on prezygotic mechanisms.

Postzygotic hybrid incompatibility can be caused by genic interactions, chromosome structure, gene transposition and reciprocal gene loss (Burke and Arnold 2001, Maheshwari and Barbash 2011). Chromosome structure can include chromosome re-arrangements, gene transposition and reciprocal gene loss. Chromosome re-arrangements can affect crossing over during meiosis and cause the production of gametes which are aneuploid or sterile (Maheshwari and Barbash 2011). Gene transposition can result in a gene being lost, especially in later generation hybrids due to random segregation (Moyle *et al.* 2010). If the gene lost codes for essential functions, the hybrid can have a reduction in fitness. Reciprocal gene loss is common in plants which have undergone whole genome duplication (WGD) as has been shown in Arabidopsis and rice (Mizuta *et al.* 2010; Bikard *et al.* 2019). In these plants, hybrid incompatibility will result when different species, sharing a common ancestor which underwent WGD, have alternative genes silenced during lineage specific evolution.

The genic mechanism underlying hybrid incompatibility follows the definition of Dobzhansky-Muller (DM) diverged genes. The DM model suggests that incompatibility is caused by the

interaction between two loci. In the ancestor, the two loci are compatible, but due to divergence during evolution of two lineages, the loci become incompatible when combined again in a hybrid (Dobzhansky 1937; Muller 1942). The interaction between the loci can either be lost or it can cause the gain of a negative interaction depending on how the orthologs have evolved. The Dobzhansky-Muller model be expanded to multiple loci but it is not yet known whether multiple loci interact together or whether the interactions are a combination of two independent interactions.

Two methods that are commonly used to identify hybrid incompatibility are QTL mapping and the analysis of segregation distortion in hybrids. QTL mapping makes use of genetic linkage maps to identify specific hybrid incompatibility traits such as hybrid sterility (Rieseberg and Carney 1998). The advantage of this approach is that a large number of loci can be identified genome-wide underlying the trait. The power of this method was seen in a recent study where four QTL were found to underlie hybrid sterility in rice (Yu *et al.* 2018). Further analysis of these regions identified two tightly linked genes, one which causes pollen abortion while the other protects pollen from abortion. When the gene which protects from pollen abortion is not present in the progeny, segregation distortion is seen in the progeny. In another study, hybrid incompatibility, in the form of hybrid weakness, was assessed in F_1 interspecific Arabidopsis hybrids (Burkart-Waco *et al.* 2012). A total of seven QTLs underlying hybrid weakness were identified, all of which were shown to interact with at least one other QTL. Further analysis of the network of QTLs identified showed that there are a large number of small effect loci which interact and control hybrid viability and growth in Arabidopsis. These studies showed that QTL mapping allows for the identification of QTLs as well as the underlying mechanism of hybrid incompatibility.

Segregation distortion is the deviation of allele frequencies from expected Mendelian ratios. Regions which show significant segregation distortion can be used to identify genes underlying hybrid incompatibility. This method allows for the whole genome to be analysed and no prior information

regarding a specific trait is required. Harushima *et al.* (2001) analysed genome-wide segregation distortion patterns in an F_2 intra-specific rice population. From the analysis, they were able to identify 33 reproductive barriers. A total of 15 of the barriers were found to affect allele transmission at the gametophyte stage while the remaining 18 affected the viability of the zygote. In another study, segregation distortion was analysed in both intra- and interspecific hybrids of oak trees (Bodénès *et al.* 2016). A total of nine significantly distorted loci were identified. The patterns of segregation distortion suggested that gametic incompatibility was a major barrier to hybridisation between oak species. The results of these studies show the power that segregation distortion analysis has to identify regions of the genome which underlie pre- and postzygotic incompatibility loci.

1.4 *Eucalyptus*

1.4.1 *Eucalyptus* domestication and hybrid breeding

Eucalyptus is planted worldwide and consists of over 700 species (Ladiges *et al.* 2003). This genus is predominantly native to Australia, with some native to neighbouring islands such as New Guinea and Timor. Due to the fast growth and wood properties of *Eucalyptus*, breeding of *Eucalyptus* increased rapidly in the 1960's (Eldridge *et al.* 1993). *Eucalyptus* is predominantly an outcrossing species (Gaiotto *et al.* 1997), therefore, when it was found that some eucalypts could undergo vegetative propagation in the 1970s, the industry grew rapidly. In plantations, *Eucalyptus* is predominantly planted as interspecific hybrids. This allows for favourable traits from different species to be combined into a single genetic background.

However, not all interspecific hybrid crosses yield successful progeny (Griffin *et al.* 1988). Hybrid incompatibility in *Eucalyptus* can occur on a pre- and postzygotic level. Not considering spatial and temporal prezygotic barriers to hybridisation, two main prezygotic barriers have been identified in *Eucalyptus*; a structural barrier and a physiological barrier to pollen tube formation. Gore *et al.* (1990) demonstrated the structural barrier of pollen tube length in *E. nitens* and *E. globulus*. They found that

the pollen tubes of *E. nitens* could not grow the entire length of the *E. globulus* style as the *E. globulus* flowers are larger than *E. nitens*. Therefore, species with different flower sizes can be difficult to cross in *Eucalyptus*. The physiological barrier occurs within the pistil where the pollen tube is inhibited resulting termination of pollen tube growth. Ellis *et al.* (1991) performed a number of intra- and inter-specific crosses as well as intergeneric crosses in *Eucalyptus*. When analysing the inter-specific crosses they found pollen-tube abnormalities in the style resulting in pre-zygotic isolation. They also found that the severity of the abnormality was correlated with the taxonomic distance of the parental species. The results of this study suggest that a prezygotic barrier in *Eucalyptus* is between the pollen and pistil and that the severity increases with an increase in taxonomic distance. These studies show the importance of both structural and physiological barriers on pre-zygotic hybrid incompatibility in *Eucalyptus*.

Postzygotic hybrid incompatibilities can occur in different stages of the trees lifecycle. Incompatibility may be seen in early stages such as a reduction in seed viability, slow germination of hybrid seed, reduced survival of germinated seed and abnormal seedling phenotype (Tibbits 1988; Lopez *et al.* 2000). Tibbits (1988) performed a number of inter- and intra-specific crosses with *E. nitens* mothers. They found that while most of the inter-specific crosses germinated, they had higher rates of abnormalities. This suggests a postzygotic barrier affecting interspecific hybrids. Lopez *et al.* (2000) performed analysis of hybrid viability over a ten year period. A number of intra- and inter-specific crosses between *E. ovata* and *E. globulus* were generated. The results showed that the inter-specific hybrids had a reduction in viability at all life stages over the ten year period. Taken together, these studies show that there is postzygotic hybrid incompatibility in *Eucalyptus* and it is important to analyse hybrid incompatibility throughout the life cycle to ensure the best performing trees are selected for further breeding.

Despite hybrid incompatibilities between *Eucalyptus*, there are many successful hybrid combinations which combine favourable traits of different species into a single genetic background. *Eucalyptus grandis* crossed with *E. urophylla* are one of the successful interspecific crosses in *Eucalyptus* (Bison *et al.* 2006). *E. grandis* is widely planted as it grows rapidly and has desirable wood properties which makes it important for pulp production (Retief and Stanger 2009), however, *E. grandis* is susceptible to many diseases (Wingfield *et al.* 1989, 1993). *E. urophylla* has been found to be more resistant than *E. grandis* to diseases (Retief and Stanger 2009), therefore, hybridisation combines the desirable pulp properties of *E. grandis* with the disease resistance of *E. urophylla* making this hybrid combination desirable for the pulp and paper industry.

1.4.2 Genetic dissection studies in *Eucalyptus*

The first genetic linkage maps for *Eucalyptus* were constructed using an F_1 biparental population, of a cross between *E. grandis* and *E. urophylla* (Grattapaglia and Sederoff 1994). The mapping population consisted of 62 interspecific hybrids and genetic maps were constructed for the *E. grandis* and *E. urophylla* parents. A total of 240 RAPD markers for *E. grandis* and 251 RAPD markers for *E. urophylla* were identified, of which only 59% and 47% respectively could be included in the genetic maps. This resulted in a low marker density of 27 cM which limited the resolution of the genetic linkage maps. For the *E. urophylla* genetic map, 11 linkage groups were identified which correspond to the 11 chromosomes in *Eucalyptus.* A total of 14 linkage groups were identified for *E. grandis* which could be due to the limited number of markers and the stringent parameters used for genetic map construction. Despite the ability to construct genetic linkage maps, this study concluded that a larger population and more markers would be needed to increase the power and accuracy of the genetic maps.

In order to identify QTLs controlling vegetative propagation in the *Eucalyptus*, the genetic maps constructed by Grattapaglia and Sederoff (1994) were re-constructed using a larger population of 112

individuals (Grattapaglia *et al.* 1995). The increase in population size resulted in 10% of the markers changing order. It was also seen that there was breakage and merging between linkage groups in the *E. grandis* genetic map, which resulted in a total of 11 linkage groups instead of 14 in the previous study. This demonstrates the important effect that sample size has on marker order and linkage group identification. A total of 20 QTLs were identified and placed on the genetic maps however, the QTLs had confidence intervals between 30-50 cM. It was therefore concluded that QTLs could only be assigned to a linkage group instead of a more precise location in the genome. To achieve more accurate locations of QTLs, the sample size and number of markers needed to be increased but this was not feasible at the time.

Many studies were conducted to improve genetic linkage maps in *Eucalyptus* after the first genetic linkage maps were constructed. The studies mainly used RAPD, AFLP and microsatellite markers. Comparisons between the studies for the same species (Table 1.1, 1.2, focus on *E. grandis* and *E. urophylla*), showed that there was a large difference in the total map length. This could be due to differences in the populations as well as the number and type of markers used. In general the earlier studies had a low marker density, low reproducibility of the markers and small samples sizes which limited the application of the genetic linkage maps. It was only once the *Eucalyptus* genome sequence became available that high density genetic linkage maps were constructed (Bartholomé *et al.* 2015).

In 2014, the *Eucalyptus* genome sequence was published (Myburg *et al.* 2014) which led to the development of a *Eucalyptus* SNP array containing 6000 markers (Bartholomé *et al.* 2015). Using the SNP array, Bartholomé *et al.* (2015), were able to genotype 1025 individuals of an *E. grandis* x *E. urophylla* interspecific cross. This resulted in high density genetic maps consisting of 2551 and 2491 markers for *E. grandis* and *E. urophylla* respectively. The average marker density was 0.36 for both *E. grandis* and *E. urophylla*, which was a significant improvement when compared with the first genetic maps constructed for *Eucalyptus.* Due to the high marker density and high resolution, the

genetic maps were used to improve the *Eucalyptus* reference genome. This study showed that in order to construct genetic linkage maps in a highly outcrossing species, with a high power and resolution, large populations and a high number of markers are required.

The above-mentioned studies focused on linkage analysis using biparental mapping populations but, association mapping in natural populations, has also been applied in *Eucalyptus*. The first association mapping study in *Eucalyptus* was a candidate gene association study, which made use of an open pollinated *Eucalyptus nitens* population, to identify alleles and haplotypes associated with microfibril angle (MFA, Thumma *et al.* 2005). The study identified two haplotypes within a gene controlling stiffness and strength in Arabidopsis. This allowed for a more detailed analysis of the region and how it controls MFA. While this study was able to identify marker-trait associations, it was not a genome-wide association study and prior information regarding the candidate region was required. Based on the findings of this study, it was concluded that for a genome-wide association study to take place, a larger number of markers would be needed.

In 2015, the first GWAS was performed in *Eucalyptus* (Cappa *et al.* 2013). A total of 303 individuals from an open-pollinated population of *Eucalyptus globulus* were genotyped using a 7,680 DArT marker array. They were able to obtain 2,364 high quality, dominant SNPs which resulted in a marker density of one marker every 260 kbp or 0.5 cM. In total 18 marker-trait associations were identified. However, this study had a low power due to the small sample size which resulted in the identification of QTL with large effects only. Therefore, it was concluded that larger sample sizes would be needed to increase the power to detect small effect QTL. Since this study there have been a limited number of GWAS in *Eucalyptus.* This is due to *Eucalyptus* being an outcrossing species and having a high rate of LD decay. Therefore, a large number of markers are required to identify all the LD blocks and until the recent development of the EUChip60K SNP chip, this was not feasible.

1.4.3 Genetic and genomic resources available for *Eucalyptus*

In 2014, the complete genome assembly of *E. grandis* was published based on the genome sequence of a 17-year-old *Eucalyptus grandis* tree produced from one generation of self-fertilization (Myburg *et al.* 2014). Sanger shotgun sequencing and paired bacterial artificial chromosome (BAC)-end sequencing were used to sequence the genome. It was estimated that 94% of the genome was assembled and the remaining 6% consisted of repeat-rich regions or regions of heterozygosity. Despite this, the assembly was largely successful and allowed for the development of new tools for *Eucalyptus* such as the *Eucalyptus* SNP chip (Silva-Junior *et al.* 2015).

The EUChip60K SNP chip was developed by Silva-Junior *et al.* (2015) and is a multispecies SNP chip. To construct the chip, 241 trees from 12 different *Eucalyptus* species were sequenced and a total of 64 639 SNPs were identified for inclusion in the chip. One concern with multispecies SNP chips is ascertainment bias, which occurs when one species is represented more often whilst making the chip. Ascertainment bias of the EUChip60K SNP chip was limited due to the data being used from many different *Eucalyptus* species. Therefore, the EUChip60K is suitable for genotyping many different Eucalypts, making it useful for GWAS and other molecular breeding strategies.

The use of multiple parents is common in hybrid breeding programmes, but have not been utilised for genetic dissection of growth and wood properties. These populations can possibly be used in a similar manner to how nested multi-parent populations have been used in crop species, as a large number of parents are crossed resulting in many F_1 hybrid progeny. Towards this, Sappi Forest Research (Hilton, KZN, South Africa) have provided access to a F_1 hybrid trial. The population consists of eight *E. grandis* pollen parents crossed with nine *E. urophylla* seed parents resulting in 17 half-sib (HS) families and 72 full-sib (FS) families. However, not all the crosses were successful and some resulted in no progeny. This provides the opportunity to analyse the parental haplotypes of successful crosses and compare them with the parental haplotypes that did not yield progeny. Marker

assisted breeding can then be used to identify parental species which are compatible and underlie desirable traits.

1.5 Conclusion

Identification of QTLs underlying quantitative traits is key for improvements in breeding programmes. Advances in mapping populations and marker identification technology have allowed for improvements in QTL identification. Multi-parent mapping populations allow for the benefits of linkage analysis and GWAS to be combined into a single population. This has allowed for high power and high resolution genetic mapping studies which have been able to identify a large number of QTLs. However, these populations have predominantly been used in species in which inbreeding can occur and have not been fully exploited in outcrossing species such as *Eucalyptus.*

Interspecific hybrid breeding allows for traits of different species to be combined in a single genetic background, which is especially useful in breeding programs. Hybrids often exhibit heterosis, resulting in them performing better than the parental species and this can be exploited in breeding programmes. However, not all hybrid combinations are compatible and this can be due to pre- or postzygotic incompatibilities. This can result in a loss of genetic diversity in hybrid breeding programmes due to incompatibility. It is therefore important to understand the mechanisms underlying hybrid incompatibility and to determine combinations of parental genotypes which are compatible.

While many studies have been performed in Eucalypts to dissect quantitative traits, these studies have limitations, resulting from lack of genomic resources, limited technology available and mapping populations which do not simultaneously have a high power and high resolution. Therefore, questions still remain regarding QTLs underlying specific traits. Despite the recent advantages in technology and techniques to analyse the quantitative traits in *Eucalyptus*, it is still not fully understood how

different combinations of QTLs in the parents combine and affect the phenotype of the progeny. With the development of the *Eucalyptus* 60K SNP chip and access to the *Eucalyptus* multi-parent mapping population (hybrid trial series), we can start to answer these questions.

The aim of this study is to dissect the genetic architecture of growth and wood properties as well as hybrid incompatibility from pure species parents into F_1 hybrid progeny. We will use a *E. grandis* HS family and a *E. urophylla* HS family from a *Eucalyptus* F_1 hybrid trial developed by Sappi Forest Research (Hilton, KZN, South Africa). Genetic linkage maps will be constructed for one *E. grandis* HS family and one *E. urophylla* HS family and QTLs controlling growth and wood properties identified. Segregation distortion of the mapped markers will be analysed as this will indicate regions of the parental genomes which underlie hybrid incompatibility. We hypothesize that there is a large amount of diversity present in the parents resulting in phenotypic variation within the F_1 progeny. We also hypothesize that there regions underlying pre- and postzygotic hybrid incompatibility between parental genomes will manifest as segregation distortion within the F_1 hybrid progeny. This study is the first step towards being able to fully utilize and exploit the advantages of multi-parent mapping populations in *Eucalyptus*.

1.6 Figures

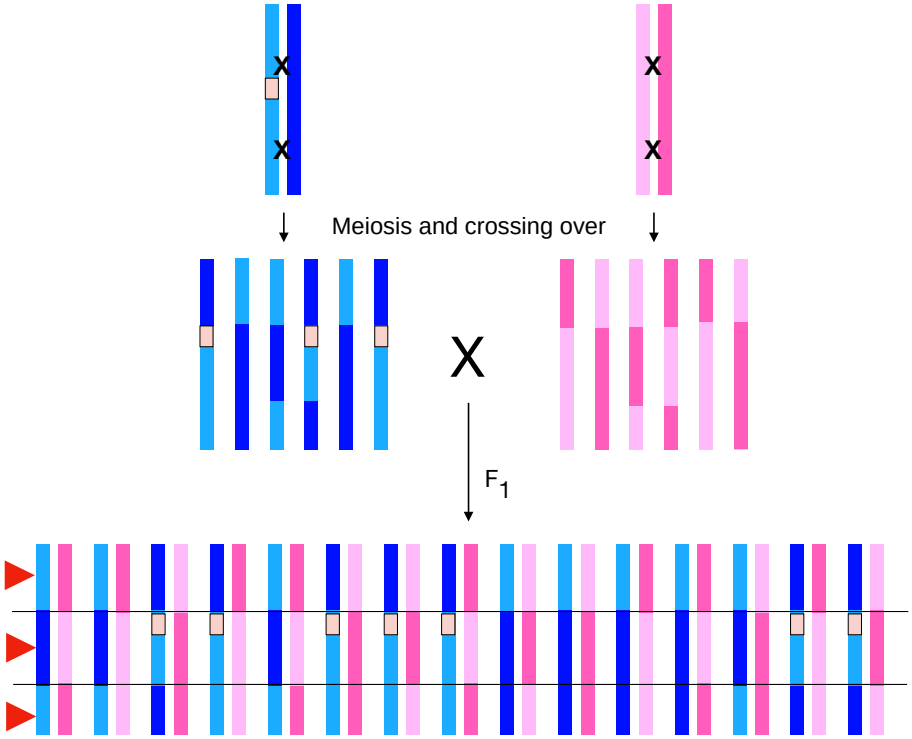


Figure 1.1 Bi-parental cross used for linkage analysis. Two parents are crossed and the F_1 progeny are analysed. Due to only one generation of mating, there is a limited amount of recombination resulting in large LD blocks (black horizonal lines). Few markers (red triangles) are required to capture all the LD blocks. A QTL (beige block) can easily be detected if it is in the same LD block as a maker as it segregates equally throughout the progeny

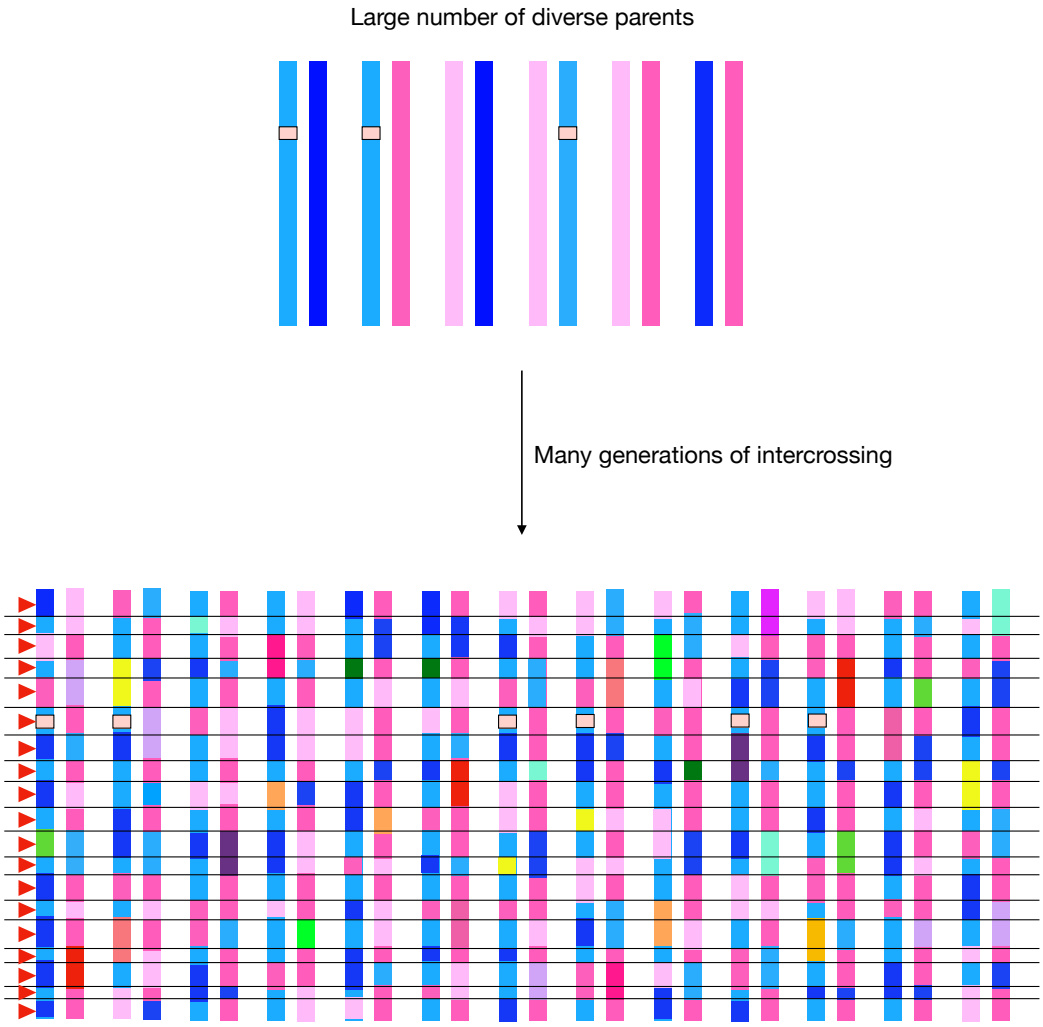


Figure 1.2 Natural population used for Genome-Wide Association Studies (GWAS). A large natural population is analysed after many generations of open pollination. This results in a large number of mutations (coloured blocks differing from pink and blue) resulting in a high allelic diversity. There are more than two alleles present resulting in rare alleles and the QTL (beige block) may only be present in a small proportion of the analysed trees. Due to many generations, the LD blocks are small (black horizontal lines) and a large number of markers (red triangles) are required to capture all of the LD blocks.

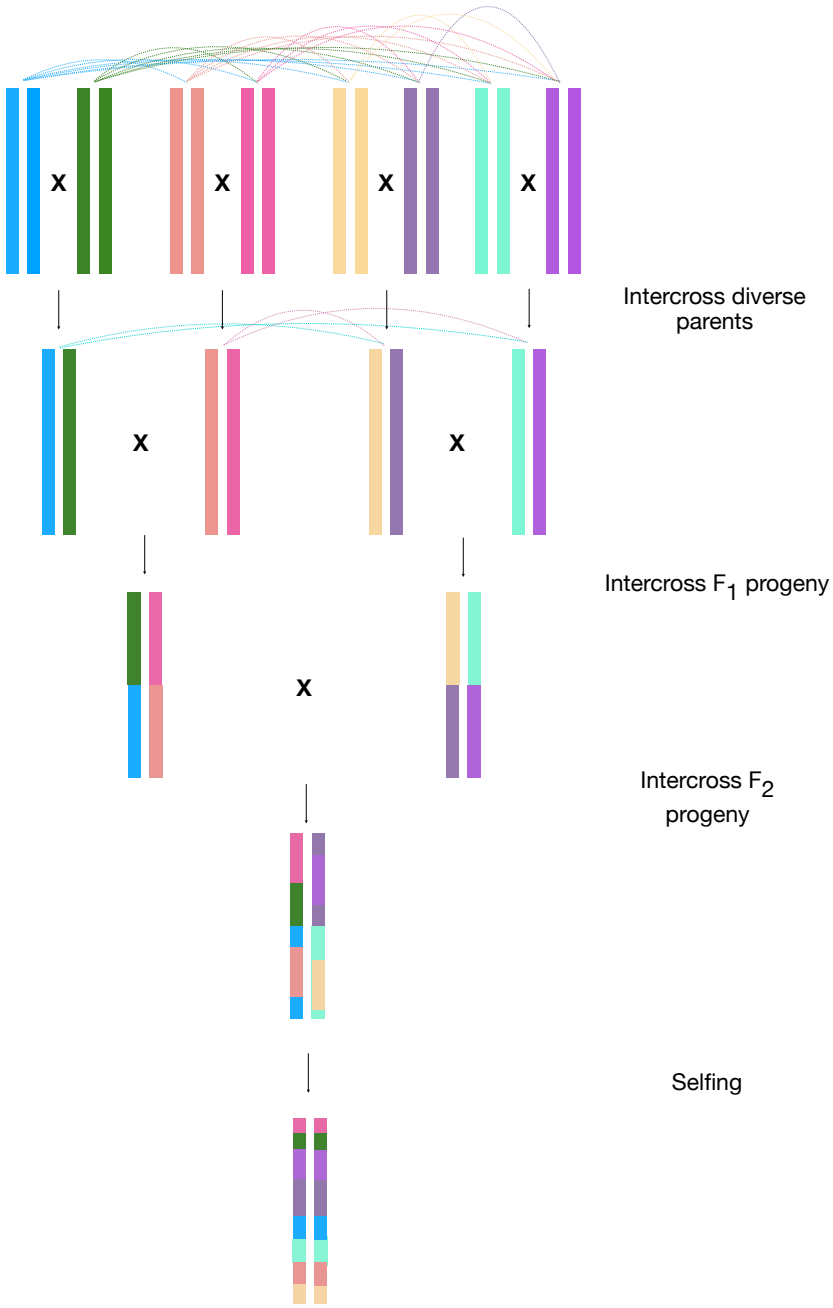


Figure 1.3 Multi-Advanced Generation Inter-Cross (MAGIC) population. Constructed by intercrossing a large number of diverse founders. Only one possible combination is shown here, dotted lines represent other possible combinations. The F_1 and F_2 progeny are intercrossed and then recombinant inbred lines are generated through selfing resulting in a high resolution. The intercrossing increases the recombination resulting in smaller LD blocks. The use of diverse founders results in a high allelic diversity with potential for rare alleles in the entire MAGIC population, but an allele will not be rare in a single RIL. Therefore this population has a high power to detect marker-trait associations.

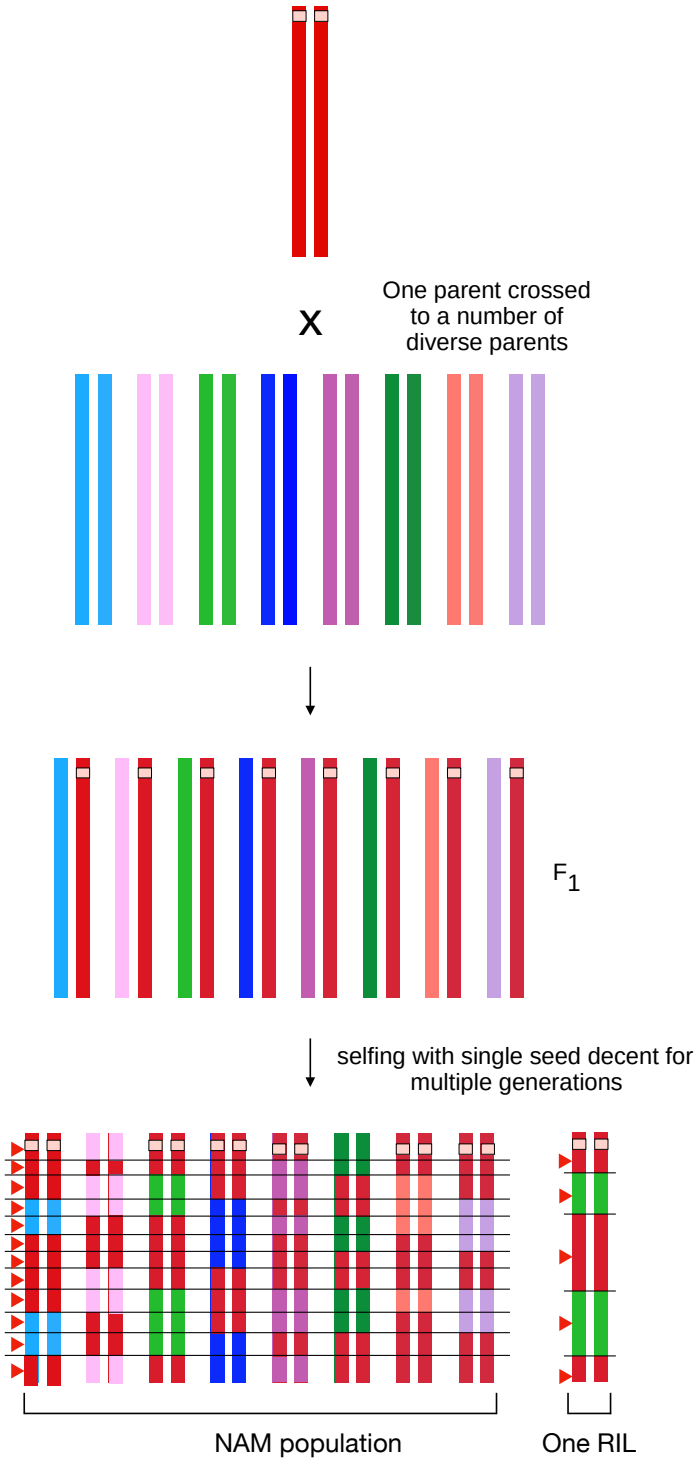


Figure 1.4 (Legend on page 24)

Figure 1.4 Nested Association Mapping (NAM) population. NAM populations are constructed by crossing a single founder to a number of diverse founders. The use of diverse founders increases the genetic diversity under study. Recombinant inbred lines are then generated which increases the amount of recombination resulting in a large number of LD blocks (black horizontal lines) within in the NAM population. This increases the resolution of the population. Within a single RIL, there are smaller LD blocks and only two alleles, resulting in a high power to detect marker-trait associations. Due to the population design, an equal number of RILs are produced, limiting the population substructure.

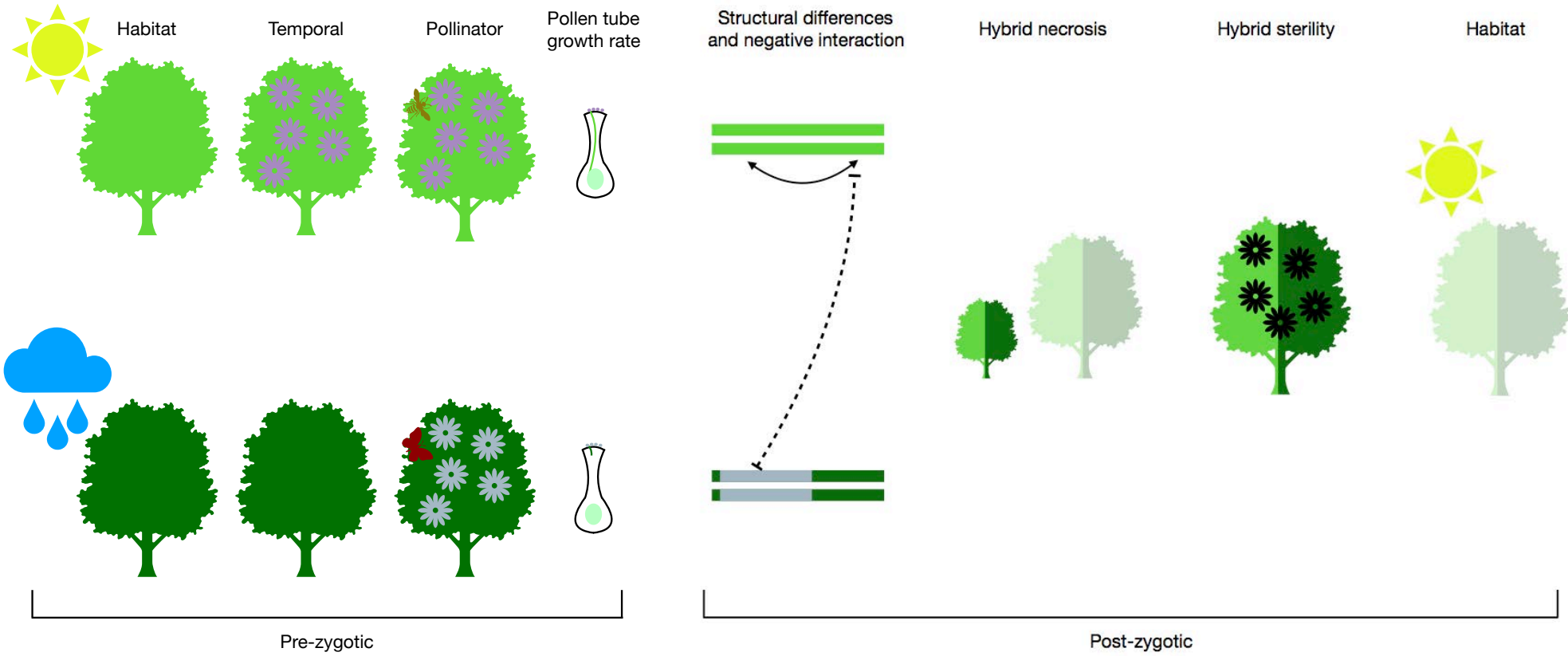


Figure 1.5 Pre- and post-zygotic barriers to hybrid compatibility (adapted from Rieseberg and Blackman 2010). Pre-zygotic mechanisms include different habitats, flowering times (temporal), pollinators and pollen tube growth rates. These are barriers which act prior to fertilisation and zygote formation. Post-zygotic barriers include structural differences and genic interactions (negative interaction) between the parental genomes. These can result in hybrid necrosis (such as stunted growth or death of hybrids), hybrid sterility and the hybrids not surviving in the habitat.

1.7 Tables

Table 1.1 Summary of genetic linkage maps constructed for *E. grandis*

Population type	Size of population	Marker type	Number of markers	Number of linkage groups	Map length (cM)	Average marker density (cM)	Reference
F_1 hybrid (GxU)	62	RAPD	240	14	1552	6.47	Grattapglia and Sederof 1994
F_1 hybrid (GxU)	93	RAPD	236	11	1415	6.00	Verhaegen *et al.* 1997
F_1 hybrid (GxU)	94	SSR (mapped to Grattapaglia and Sederoff 1994)	20	11	Mapped to Grattapaglia and Sederof 1994)	NA	Brondani *et al.* 1998
F_1 hybrid (GxU)	92	SSR (mapped to Brondani *et al*. 1998)	50 new SSR	11	2088	33.14	Brondani *et al.* 2002
E. grandis BC	156	AFLP	138	11	1335	9.67	Myburg *et al.* 2003
F_1 hybrid (GxU)	92	SSR	202	11	1814.5	8.98	Brondani *et al.* 2006
E. grandis BC	180	DArT and SSR	957 DArT, 34 SSR	11	924.7	0.93	Kullan *et al.* 2012a
F_1 hybrid (GxU)	1025	SNP	2551	11	912.59	0.36	Bartholome *et al*. 2015

Table 1.2 Summary of genetic linkage maps constructed for *E. urophylla*

Population type	Size of population	Marker type	Number of markers	Number of linkage groups	Map length (cM)	Average marker density (cM)	Reference
F_1 hybrid (GxU)	62	RAPD	251	11	1101	4.39	Grattapglia and Sederof 1994
F_1 hybrid (GxU)	93	RAPD	269	11	1331	4.95	Verhaegen and Plomion 1996
F_1 hybrid (GxU)	94	SSR	20	11	Mapped to Grattapaglia and Sederof 1994)	NA	Brondani *et al.* 1998
F_1 hybrid (GxU)	92	SSR (Mapped to Brondani *et al.* 1998)	50 new SSR	10	1804	20.45	Brondani *et al.* 2002
F_1 hybrid (UxT)	82	RAPD	220	23	1504.6	6.84	Gan *et al.* 2003
F_1 hybrid (GxU)	92	SSR	160	11	1133.4	7.08	Brondani *et al.* 2006
E. urophylla BC	367	DArT and SSR	912 DArT, 46 SSR	11	1107.3	1.16	Kullan *et al.* 2012a
F_1 hybrid (GxU)	1025	SNP	2491	11	903.99	0.36	Bartholome *et al.* 2015

1.8 References

Abiola O, Angel JM, Avner P, Bachmanov AA, Belknap JK, Bennet B, Blankenhorn EP, Blizard DA, Bolivar V, Brockmann GA, *et al.* 2003. The nature and identification of quantitative trait loci : a community's view. *Nature Reviews Genetics* 4: 911–916.

Bajgain P, Rouse MN, Tsilo TJ, Macharia GK, Bhavani S, Jin Y, Anderson JA. 2016. Nested association mapping of stem rust resistance in wheat using genotyping by sequencing. *PLoS ONE* 11: e0155760.

Bandillo N, Raghavan C, Muyco PA, Sevilla MAL, Lobina IT, Dilla-Ermita CJ, Tung C-W, McCouch S, Thomson M, Mauleon R, *et al.* 2013. Multi-parent advanced generation inter-cross (MAGIC) populations in rice: progress and potential for genetics research and breeding. *Rice* 6: 11.

Bartholomé J, Mandrou E, Mabiala A, Jenkins J, Nabihoudine I, Klopp C, Schmutz J, Plomion C, Gion J. 2015. High-resolution genetic maps of *Eucalyptus* improve *Eucalyptus grandis* genome assembly. *New Phytologist* 206: 1283–1296.

Bikard D, Patel D, Mette C Le, Giorgi V, Bennett MJ, Loudet O. 2019. Genes leads to genetic divergent evolution of duplicate incompatibilities within *A. thaliana*. *Science* 323: 623–626.

Bison O, Ramalho MAP, Rezende GDSP, Aguiar AM, de Resende MDV. 2006. Comparison between open pollinated progenies and hybrids performance in *Eucalyptus grandis* and *Eucalyptus urophylla*. *Silvae Genetica* 55: 192–196.

Bodénès C, Chancerel E, Ehrenmann F, Kremer A, Plomion C. 2016. High-density linkage mapping and distribution of segregation distortion regions in the oak genome. *DNA Research* 23: 115–124.

Bouchet S, Olatoye MO, Marla SR, Perumal R, Tesso T. 2017. Increased power to dissect adaptive traits in global sorghum diversity using a nested association mapping population. *Genetics* 206: 573–585.

Brondani R, Brondani C, Grattapaglia D. 2002. Towards a genus-wide reference linkage map for *Eucalyptus* based exclusively on highly informative microsatellite markers. *Molecular Genetics and Genomics* 267: 338–347.

Brondani RP V, Brondani C, Tarchini R, Grattapaglia D. 1998. Development, characterization and mapping of microsatellite markers in *Eucalyptus grandis* and *E. urophylla*. *Theoretical and Applied Genetics* 97: 816–827.

Buckler ES, Holland JB, Bradbury PJ, Acharya CB, Brown PJ, Browne C, Ersoz E, Flint-Garcia S, Garcia A, Glaubitz JC, *et al.* 2009. The genetic architecture of maize flowering time. *Science* 325: 714–718.

Burkart-Waco D, Josefsson C, Dilkes B, Kozloff N, Torjek O, Meyer R, Altmann T, Comai L. 2012.

Hybrid incompatibility in Arabidopsis is determined by a multiple-locus genetic network. *Plant Physiology* 158: 801–812.

Burke JM, Arnold ML. 2001. Genetics and the fitness of hybrids. *Annual Review of Genetics* 35: 31–52.

Cappa EP, El-Kassaby YA, Garcia MN, Acuña C, Borralho NMG, Grattapaglia D, Marcucci Poltri SN. 2013. Impacts of population structure and analytical models in genome-wide association studies of complex traits in forest trees : A case study in *Eucalyptus globulus*. *PLoS ONE* 8: e81267.

Collard BCY, Jahufer MZZ, Brouwer JB, Pang ECK. 2005. An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: The basic concepts. *Euphytica* 142: 169–196.

Darvasi A, Soller M. 1995. Advanced intercross lines, an experimental population for fine genetic mapping. *Genetics* 141: 1199–1207.

Dell'Acqua M, Gatti DM, Pea G, Cattonaro F, Coppens F, Magris G, Hlaing AL, Aung HH, Nelissen H, Baute J, *et al.* 2015. Genetic properties of the MAGIC maize population: A new platform for high definition QTL mapping in *Zea mays*. *Genome Biology* 16: 1–23.

Dobzhansky T. 1937. *Genetics and the origin of species*. New York: Columbia University Press.
Eldridge K, Davidson J, Hardwood C, van Wyk G. 1993. *Eucalypt domesitcation and breeding*. Oxford: Oxford University Press.

Ellis MF, Sedgley M, Gardner JA. 1991. Interspecific pollen-pistil interaction in *Eucalyptus* L ' Hér. (Myrtaceae): The effect of taxonomic distance. *Annals of Botany* 68: 185–194.

Flint-Garcia SA, Thornsberry JM, Buckler ES. 2003. Structure of linkage disequilibrium in plants. *Annual Review of Plant Biology* 54: 357–74.

Fragoso CA, Moreno M, Wang Z, Heffelfinger C, Arbelaez LJ, Aguirre JA, Franco N, Romero LE, Labadie K, Zhao H, *et al.* 2017. Genetic architecture of a rice nested association mapping population. *G3* 7: 1913–1926.

Gaiotto FA, Bramucci M, Grattapaglia D. 1997. Estimation of outcrossing rate in a breeding population of *Eucalyptus urophylla* with dominant RAPD and AFLP markers. *Theoretical and Applied Genetics* 95: 842–849.

Gore PL, Potts BM, Volker PW, Megalos J. 1990. Unilateral cross-incompatibility in *Eucalyptus*: The case of hybridisation between *E . globulus* and *E . nitens*. *Australian Journal of Botany* 38: 383–394.

Grattapaglia D, Bertolucci FL, Sederoff RR. 1995. Genetic mapping of QTLs controlling vegetative propagation in *Eucalyptus grandis* and *E. urophylla* using a pseudo-testcross strategy and RAPD

markers. *Theoretical and Applied Genetics* 90: 933–947.

Grattapaglia D, Sederoff R. 1994. Genetic linkage maps of *Eucalyptus grandis* and *Eucalyptus urophylla* using a pseudo-testcross: Mapping strategy and RAPD markers. *Genetics* 137: 1121–1137.

Griffin AR, Burgess IP, Wolf L. 1988. Patterns of natural and manipulated hybridisation in the genus *Eucalyptus* L'Herit - a review. *Australian Journal of Botany* 107: 41–66.

Harushima Y, Nakagahra M, Yano M, Sasaki T, Kurata N. 2001. A genome-wide survey of reproductive barriers in an intraspecific hybrid. *Genetics* 159: 883–892.

Hirschhorn JN, Daly MJ. 2005. Genome-wide association studies for common diseases and complex traits. *Nature Reviews Genetics* 6: 95–108.

Huang X, Zhao Y, Wei X, Li C, Wang A, Zhao Q, Li W, Guo Y, Deng L, Zhu C, *et al.* 2012. Genome-wide association study of flowering time and grain yield traits in a worldwide collection of rice germplasm. *Nature Genetics* 44: 32–39.

Kover PX, Valdar W, Trakalo J, Scarcelli N, Ehrenreich IM, Purugganan MD, Durrant C, Mott R. 2009. A multiparent advanced generation inter-cross to fine-map quantitative traits in *Arabidopsis thaliana*. *PLoS Genetics* 5: e1000551.

Ladiges PY, Udovicic F, Nelson G. 2003. Australian biogeographical connections and the phylogeny of large genera in the plant family Myrtaceae. *Journal of Biogeography* 30: 989–998.

Lander ES, Schork NJ. 1994. Genetic dissection of complex traits. *Science* 265: 2037–2048.

Lopez GA, Potts BM, Tilyard PA. 2000. F1 hybrid inviability in *Eucalyptus* : the case of *E . ovata* x *E . globulus*. *Heredity* 85: 242–250.

Mackay TF. 2001. The genetic architecture of quantitative traits. *Annual Review of Genetics* 35: 303–339.

Mackay TFC, Stone EA, Ayroles JF. 2009. The genetics of quantitative traits: Challenges and prospects. *Nature Reviews Genetics* 10: 565–577.

Maheshwari S, Barbash DA. 2011. The genetics of hybrid incompatibilities. *Annual Review of Genetics* 45: 331–355.

Maurer A, Draba V, Jiang Y, Schnaithmann F, Sharma R, Schumann E, Kilian B, Reif JC, Pillen K. 2015. Modelling the genetic architecture of flowering time control in barley through nested association mapping. *BMC Genomics* 16: 290.

McMullen MD, Kresovich S, Villeda HS, Bradbury P, Li H, Sun Q, Flint-garcia S, Thornsberry J, Acharya C, Bottoms C, *et al.* 2009. Genetic properties of the maize nested association mapping

population. *Science* 325: 737–740.

Mizuta Y, Harushima Y, Kurata N. 2010. Rice pollen hybrid incompatibility caused by reciprocal gene loss of duplicated genes. *Proceedings of the National Academy of Sciences* 107: 20417–20422.

Moyle LC, Muir CD, Han M V, Hahn MW. 2010. The contribution of gene movement to the 'two rules of speciation'. *Evolution* 64: 1541–1557.

Muller HJ. 1942. Isolating mechanisms, evolution and temperature. *Biology Symposium.* 6: 71–125.

Myburg AA, Grattapaglia D, Tuskan GA, Hellsten U, Hayes RD, Grimwood J, Jenkins J, Lindquist E, Tice H, Bauer D, *et al.* 2014. The genome of *Eucalyptus grandis*. *Nature* 510: 356–362.

Pascual L, Desplat N, Huang BE, Desgroux A, Bruguier L, Bouchet JP, Le QH, Chauchard B, Verschave P, Causse M. 2015. Potential of a tomato MAGIC population to decipher the genetic control of quantitative traits and detect causal variants in the resequencing era. *Plant Biotechnology Journal* 13: 565–577.

Pierce BA. 2014. *Genetics : a conceptual approach*. New York: W.H. Freeman and Company.

Retief ECL, Stanger TK. 2009. Genetic parameters of pure and hybrid populations of *Eucalyptus grandis* and *E. urophylla* and implications for hybrid breeding strategy. *Southern Forests: a Journal of Forest Science* 71: 133–140.

Rieseberg LH, Blackman BK. 2010. Speciation genes in plants. *Annals of Botany* 106: 439–455.

Rieseberg LH, Carney SE. 1998. Plant hybridization. *New Phytologist* 140: 599–624.

Risch N, Merikangas K. 1996. The future of genetic studies of complex human diseases. *Science* 273: 1516–1517.

Sannemann W, Huang BE, Mathew B, Léon J. 2015. Multi-parent advanced generation inter-cross in barley: high-resolution quantitative trait locus mapping for flowering time as a proof of concept. *Molecular Breeding* 35: 86.

Silva-Junior OB, Faria DA, Grattapaglia D. 2015. A flexible multi-species genome-wide 60K SNP chip developed from pooled resequencing of 240 *Eucalyptus* tree genomes across 12 species. *New Phytologist* 206: 1527–1540.

Snow AA, Spira TP, Liu H. 2000. Effects of sequential pollination on the success of 'fast' and 'slow' pollen donors in *Hibiscus moscheutos* (Malvaceae). *American Journal of Botany* 87: 1656–1659.

Song Q, Yan L, Quigley C, Jordan BD, Fickus E, Schroeder S, Song B-H, Charles An Y-Q, Hyten D, Nelson R, *et al.* 2017. Genetic characterization of the soybean nested association mapping population. *The Plant Genome* 10: 2

Tang J, Yan J, Ma X, Teng W, Wu W, Dai J, Dhillon BS, Melchinger AE, Li J. 2010. Dissection of the genetic basis of heterosis in an elite maize hybrid by QTL mapping in an immortalized F2 population. *Theoretical and Applied Genetics* 120: 333–340.

Thumma BR, Nolan MF, Evans R, Moran GF. 2005. Polymorphisms in cinnamoyl CoA reductase (CCR) are associated with variation in microfibril angle in *Eucalyptus* spp. *Genetics* 171: 1257–1265.

Tibbits WN. 1988. Germination and morphology of progeny from controlled pollinations of *Eucalyptus nitens* (Deane & Maiden) Maiden. *Australian Journal of Botany* 36: 677–691.

Verhaegen D, Plomion C, Gion JM, Poitel M, Costa P, Kremer A. 1997. Quantitative trait dissection analysis in *Eucalyptus* using RADP markers: 1. Detection of QTL in interspecific hybrid progeny, stability of QTL expression across different ages. *Theoretical and Applied Genetics* 95: 597–608.

Wingfield MJ, Crous IPW, Swart WJ. 1993. *Sporothrix eucalypti* (sp. nov.), a shoot and leaf pathogen of *Eucalyptus* in South Africa. *Mycopathologia* 123: 159–164.

Wingfield MJ, Swart WJ, Abear BJ. 1989. First record of *Cryphonectria* canker of *Eucalyptus* in South Africa. *Phytophylactica* 21: 311–313.

Yu J, Holland JB, McMullen MD, Buckler ES. 2008. Genetic design and statistical power of nested association mapping in maize. *Genetics* 178: 539–551.

Yu X, Zhao Z, Zheng X, Zhou J, Kong W, Wang P, Bai W, Zheng H, Zhang H, Li J, *et al.* 2018. A selfish genetic element confers non-Mendelian inheritance in rice. *Science* 360: 1130–1132.

Zhou G, Chen Y, Yao W, Zhang C, Xie W, Hua J, Xing Y, Xiao J, Zhang Q. 2012. Genetic composition of yield heterosis in an elite rice hybrid. *Proceedings of the National Academy of Sciences*: 1–6.

Zhu C, Gore M, Buckler ES, Yu J. 2008. Status and prospects of association mapping in plants. *The Plant Genome Journal* 1: 5-20.

Chapter 2

Identification of QTL underlying growth and wood properties in a nested, multi-parent *Eucalyptus* hybrid population

Julia Candotti[1], Marja M. O'Neill[1], S. Melissa Reynolds[1], Roobavathie Naidoo[2], Nicoletta Jones[2], Eshchar Mizrachi[1], Alexander A. Myburg[1*]

[1] *Department of Biochemistry, Genetics and Microbiology, Forestry and Agricultural Biotechnology Institute (FABI), University of Pretoria, Private bag X20, Pretoria 0028, South Africa*

[2] *Sappi Forests Research, Shaw Research Centre, PO Box 473, Howick, 3290, South Africa*

This research chapter has been prepared in the format required for submission to a peer-reviewed journal (*Tree Genetics and Genomes*). I performed all analyses in this manuscript and prepared the manuscript. Mrs M.M. O'Neill and Ms S.M. Reynolds provided technical support and advise on data analysis throughout the project. Ms R. Naidoo and Dr N. Jones constructed and maintained the multi-parent population, provided all sample tissue and performed trait measurements. Prof E. Mizrachi co-supervised the project. Prof A.A. Myburg conceived and supervised the project as well as provided valuable revisions for the manuscript.

2.1 Abstract

As demonstrated in crop species, multi-parent mapping approaches can provide increased power and resolution for identifying marker-trait associations compared to genome-wide association studies (GWAS) and linkage (quantitative trait loci (QTL)) analysis. This strategy has not been fully exploited in outcrossing forest tree species. As a test of feasibility, we performed genetic dissection of growth and wood traits in a nested, multi-parent *Eucalyptus grandis* x *E. urophylla* F_1 hybrid trial. The population was constructed by crossing nine *E. grandis* pollen parents with eight *E. urophylla* seed parents and planted across four sites. From this population, one *E. grandis* half-sib (HS) family and one (intersecting) *E. urophylla* HS family were used in this study. A total of 349 and 367 individuals for the *E. grandis* and *E. urophylla* HS families, respectively, were genotyped using the EUChip60K SNP chip. A total of 2124 and 2015 informative single nucleotide polymorphisms (SNP) markers were identified for the *E. grandis* and *E. urophylla* HS families, respectively. The markers were used to construct framework genetic linkage maps for each parent onto which QTLs were mapped. A total of 15 and 23 QTLs underlying growth and wood properties were identified across the four sites for the *E. grandis* and *E. urophylla* HS families respectively. The percentage of phenotypic variance explained by the QTLs ranged from 3.06% and 36.58%. Genotype-by-environment interaction possibly affected trait expression as different QTLs were identified in the sites for most traits. This study represents an important first step towards genetic dissection of complex trait variation in *E. grandis* x *E. urophylla* multi-parent F_1 hybrid trials.

2.2 Introduction

Quantitative genetics is important for understanding the relationship between genotype and phenotype. Complex traits are affected by many QTL each of which have varying effects on the phenotype and is modified by the environment. It is important to identify markers which are associated with QTLs so that the markers can be used for marker assisted breeding (MAB) in breeding programmes. MAB is especially important in species which have long generation times as it allows

for early selection as well as the monitoring of the genetic diversity present in breeding programmes. Marker-trait associations are typically identified using linkage analysis or GWAS.

Linkage analysis traditionally makes use of bi-parental crosses to identify markers linked to a trait of interest (Mackay 2001). Due to only two alleles of each parent segregating in the population, linkage analysis has high statistical power to detect marker-trait associations. However, due to the limited amount of recombination, linkage disequilibrium (LD) blocks are large which limits the resolution of linkage analysis in biparental crosses. GWAS typically use large, natural populations to identify markers associated with a trait of interest. This is advantageous as there is a large amount of historical recombination in natural populations, resulting in small LD blocks which gives GWAS a high resolution and highly variable allele frequencies (Hirschhorn and Daly 2005). However, due to the large amount of allelic diversity segregating in natural populations, the power to detect marker-trait associations, especially for rare alleles, is limited.

Multi-parent mapping approaches aim to combine the high power of linkage analysis with the high resolution of GWAS. Two commonly used multi-parent population types are nested association mapping (NAM, Yu *et al.* 2008) and multi-parent advanced generation intercross (MAGIC, Kover *et al.* 2009). These populations have been constructed by crossing a number of diverse founders followed by the generation of recombinant inbred lines (RIL). A number of studies have used such approaches in crop species and have been able to identify larger number of QTLs of varying effects when compared with previous population types. For example, in the first Arabidopsis MAGIC population, QTLs were identified with a higher precision and resolution for germination data and bolting time when compared with previous studies (Kover *et al.* 2009). However, these mapping approaches have been limited to species in which RILs can be developed and the approach has not been fully utilised in outcrossing species.

In order for marker assisted selection (MAS) to be successful, the environment also needs to be taken into consideration as MAS has been shown to be affected by the environment. Peng-yuan *et al.* (2006) used models to determine the effect of the environment on MAS. They found that when one environment was used to identify markers for MAS, the application of these markers had a reduced success. However, when QTLs were identified across multiple sites, MAS was more successful. This is due to genotypes interacting differently with each environment through genotype-by-environment (GxE) interaction. Therefore, in order to apply MAS to breeding programmes, it is valuable to use multiple environments to develop marker-trait associations, especially if MAS will be used across multiple environments.

Eucalyptus is an outcrossing genus which is planted globally in tropical, subtropical and some temperate regions. *Eucalyptus* is often planted as interspecific F_1 hybrid clones as this allows for favourable traits from two species to be combined in a single genetic background (de Assis 2000). One of the most commonly planted interspecific hybrid combinations in subtropical sites is *Eucalyptus grandis* x *E. urophylla*. *E. grandis* has desirable growth and wood properties (Retief and Stanger 2009), however, it is highly susceptible to disease (Wingfield *et al.* 1989). Therefore, *E. grandis* is crossed with *E. urophylla* which is more disease resistant (Retief and Stanger 2009).

The *Eucalyptus* genome was published in 2014 (Myburg *et al.* 2014) which enabled the development of the *Eucalyptus* EUChip60K SNP chip (Silva-Junior *et al.* 2015). The SNP chip provides a high throughput genotyping platform which enables large populations to be genotyped. This has enabled advances in population genomic studies in *Eucalyptus* (Grattapaglia *et al.* 2018). *Eucalyptus* hybrid breeding programmes generate F_1 hybrid trials by crossing a large number of diverse individuals. However, these populations have not been explored as possible multi-parent mapping populations for genetic dissection of quantitative traits. Recently, a *Eucalyptus* multi-parent mapping population was obtained from a F_1 hybrid trial constructed by Sappi Forest Research (Hilton, KZN, South Africa).

The population was constructed by crossing nine *E. grandis* pollen parents with eight *E. urophylla* seed parents. The population was planted across four different sites which enables QTL analysis within and across environments.

This study focused on one *E. grandis* HS family and one *E. urophylla* HS family of the *Eucalyptus* F_1 hybrid breeding trial. The aim of this study was to construct framework genetic linkage maps for the *E. grandis* pollen parent and the *E. urophylla* seed parent and to map QTLs underlying growth and wood properties in the two HS families. This study represents an important first step towards genetic dissection of hybrid combining ability and complex trait variation in *E. grandis* x *E. urophylla* multi-parent F_1 hybrid trials.

2.3 Methods

2.3.1 Plant material and DNA isolation

A *Eucalyptus* multi-parent population was constructed by crossing nine *E. grandis* pollen parents with eight *E. urophylla* seed parents to generate F_1 hybrid progeny (Sappi Forest Research, South Africa, Table 2.1). This study focused on one *E. grandis* HS family and one *E. urophylla* HS family (Table 2.1). The population was planted across four different sites in South Africa (Supplementary Table 2.1). DNA was isolated from leaf and wood tissue using the NucleoSpin® Plant II DNA extraction kit (Machery-Nagel, Germany) with some modifications to the protocol.

2.3.2 Trait data

A total of six traits were analysed at four years of age; diameter at breast height (DBH), height, volume, wood density, near-infrared range (NIR) dissolving pulp yield (dPY) and NIR S:G ratio (S:G). A two-way ANOVA was performed for the trait data for each HS family with full-sib (FS) family and site being the two variables analysed. A one-way ANOVA was performed on each site with FS family as a variable. Data was standardized across the sites ((mean of site – individual value)

/ standard deviation of site). A two-way and one-way ANOVA was performed on the corrected data for each HS and each site. Data was analysed for normality using the Shapiro-Wilk. Correlations between the different traits were analysed on the corrected data using both Pearson's and Spearman's correlation.

2.3.3 Parentage analysis

To confirm the parents of the individuals of the *E. grandis* HS family, 10 microsatellite markers were used. Microsatellite markers were PCR amplified using the Qiagen® Multiplex PCR kit (Qiagen, MD, USA). GeneScan™ fragment length analysis was performed at the University of Pretoria DNA Sequencing Facility, using the LIZ™ 500 size standard, G5 filter set and an ABI3500XL DNA sequencer (Applied Biosystems, CA, USA). The allele sizes of the microsatellite fragments were analysed using GeneMarker® 1.95 (Softgenetics, State College, PA, USA) and the results were exported to Excel for further analyses.

Discrete allele matching was used to determine if the allele sizes of the seedlings matched that of the expected parent. The data was further analysed in CERVUS v 3.0.7 (Kalinowski *et al.* 2007), using the default settings, to confirm parentage and to identify alternative parents where the progeny mismatched parents. COLONY v2.0.6.4 (Jones and Wang 2010) was used to determine pedigree structure as it identifies HS and full-sib (FS) relationships (Supplementary Table 2.2).

Parentage analysis for the *E. urophylla* HS family was performed with SNP markers using identity by decent (IBD) analysis. Markers with a HWE $P > 1^{e-0.5}$ were removed and the Identity by Decent (IBD) Estimation function in SVS 8.7.1 (SVS, Golden Helix®, Inc. Bozeman, MT) was used. The IBD estimation was calculated using: Output PI = P(Z=1)/2 +P(Z=2) and all pairs where the PI was >= 0 were recorded.

2.3.4 SNP genotyping

Samples were genotyped using the *Eucalyptus* EUChip60K SNP chip (Silva-Junior *et al.* 2015) at GeneSeek® (Neogen, Lincoln, NE USA). Genotypic classes were identified using GenomeStudio® 2.0 (Illumina, CA, USA) as described by Silva-Junior et al. (2015). Briefly, for each HS family, 70 samples were selected with each FS family represented equally. Samples and SNPs which met the following criteria were used to identify the cluster positions and a cluster position file was generated: (a) samples with a call rate > 0.8, (b) SNPs with cluster separation values > 0.3, (c) SNPs with mean normalized intensity values for the heterozygous cluster > 0.2, (d) SNPs with mean normalized theta value for the heterozygous cluster > 0.2 and < 0.8, (e) removed markers which showed deviation from the expected inheritance patterns (Parent-Parent-Child (P-P-C) > 50). Cluster position files were generated and applied to the entire HS families data.

Following reclustering, a final report was generated and quality control performed in Golden Helix SVS 8.7.1 (SVS8, Golden Helix®, Inc., Bozeman, MT). A physical map was applied to the genotypic data for all markers with unique positions in the version 2.0 *Eucalyptus grandis* genome assembly (https://phytozome.jgi.doe.gov/). SNPs with a call rate < 0.9 and a minor allele frequency (MAF) < 0.05 were removed. Samples with a call rate < 0.9 were excluded.

2.3.5 Species discrimination

A principal component analysis (PCA) was performed using the full set of SNP data in order to confirm the species and hybrid status of the FS families. Prior to the PCA, markers were retained if they met the following criteria in SVS 8.7.1: call rate > 0.9 and MAF > 0.05. The Genotype Principal Component Analysis function in SVS 8.7.1 was used to construct the PCA for each HS family. The following parameters were selected: number of components was set to the maximum number of subpopulations expected, additive genetic model selected and each marker was normalised by its theoretical standard deviation under Hardy-Weinberg Equilibrium (HWE). The results of the analysis

were visualised by plotting the two highest Eigenvalues as a scatter plot. Reference datasets (Reynolds *et al.* 2019 unpublished) for *E. grandis, E. nitens, E. urophylla* and *E. dunnii* were included in this analysis.

2.3.6 Identification of informative SNPs

Informative markers were markers with the following criteria: markers heterozygous in the common parent (MAF > 0.4) and homozygous in the other parents (MAF < 0.01). SNP markers were removed from the mapping dataset if two homozygous classes were present in the progeny. SNP markers which were segregating within 90% of the progeny of each FS family and markers which had a 90% call rate in each FS family were retained. SNP markers were removed if more than two FS family parents had a missing genotype for that marker or if two homozygous classes were present in the FS parents.

2.3.7 Genetic map construction

Genetic maps were constructed in JoinMap® 4.1 (Van Ooijen 2006) using a pseudo-test-cross mapping strategy (Grattapaglia and Sederoff 1994). The informative markers were coded according to the segregation class <nnxnp>, with an expected segregation ratio of 1:1. A cross pollinated (CP) population type was selected and a logarithm-of-the-odds was used to define the linkage groups. A recombination frequency of 0.4 and regression mapping using Kosambi's mapping function with the default parameters was used to order the loci.

Identical loci were removed in JoinMap using the "Remove Identicals" function under the population node. The χ^2 test in JoinMap 4.1 was used to evaluate Mendelian segregation rates. After each round of mapping the markers were analysed using the genotype probability function and markers with a Nearest Neighbour fit (N.N fit) > 3 cM were removed. Markers were then re-ordered and the process continued until all markers met the parameters. Full genetic linkage maps were constructed to contain

the maximum number of SNP markers possible. The R package LinkageMapView 2.1.2 (Ouellette *et al.* 2018) was used to visualise and compare the collinearity of the genetic linkage maps with the physical map based on version 2 of the *Eucalyptus grandis* genome assembly (https://phytozome.jgi.doe.gov/).

Framework genetic linkage maps were subsequently constructed with the aim to achieve an average marker spacing of 2 cM to facilitate QTL detection. The same parameters used for the construction of the high density genetic linkage maps were used here, with additional criteria were that marker intervals should not be smaller than 1 cM or larger than 10 cM. The framework genetic linkage map and the physical position map were visualised using LinkageMapView 2.1.2 (Ouellette *et al.* 2018). The coverage of the framework genetic linkage maps were calculated using the formula from (Lange and Boehnke 1982).

2.3.8 QTL mapping

QTL Cartographer v1.17 (Basten *et al.* 1994, 2004) was used to identify QTLs. A backcross (BC1) population model type was used by converting markers scored nn to 1 and np to 0. The likelihood ratio test statistic (LR) threshold was determined using a composite interval permutation test of 1000 permutations. Composite interval mapping (Zeng 1993, 1994) was performed with forward and backward elimination (P-value = 0.1) and a walking speed of 1 cM. The HS families were analysed separately and each HS was analysed across all sites with more than 80 individuals. Other traits (o-traits) included in the QTL model were site (only when mapping in HS families) and family (when mapping in HS families and across the sites). QTL peaks which were more than 20 cM apart were considered as separate QTLs. QTL peak profiles were visualised in R and the QTLs were visualised on the framework genetic linkage maps using LinkageMapView 2.1.2 (Ouellette *et al.* 2018).

For identified growth QTLs in significantly distorted regions, the marker closest to the peak was identified. The trait data was separated into two groups based on the genotypic classes of the marker (0,1). The mean trait value was calculated and visualised using a dot and box plot. The number of individuals in each genotypic class was visualised using a bar plot.

2.4 Results

2.4.1 Parentage confirmation

We used microsatellite markers to confirm the parentage of 591 seedlings from the *E. grandis* HS family. We found that the seed parent of one FS family was not the seed parent, but the 96 progeny were shown to be full-siblings matching the pollen parent and therefore retained in the mapping dataset. In total, we were able to confirm the parentage of 518 samples for the *E. grandis* HS family. For the *E. urophylla* HS family we genotyped 393 seedlings. Following IBD analysis, we found that 24 samples did not belong to FS families in this study and four samples were re-assigned to a different family. A total of 369 individuals in the *E. urophylla* HS family had parentage confirmed.

2.4.2 Species identification

We performed a PCA to confirm that the individuals of this population are *E. grandis* x *E. urophylla* (GU) hybrids. From the PCA (Supplementary Figure 2.1), we confirmed that 12 of the FS families were *E. grandis* x *E. urophylla* hybrids (five FS families for the *E. grandis* HS family and seven FS families for the *E. urophylla* HS family). We found that one *E. urophylla* seed parent was a *E. grandis x E. urophylla* F_1 hybrid and we removed the seedlings of this parent (F_2 backcross) from the study. We observed that the seedlings from the FS family, for which we could not identify the seed parent using microsatellite markers, clustered with *E. grandis* references. These results show that the seed parent was most likely *E. grandis* individual instead of *E. urophylla* and we removed the samples from the study. Altogether, we confirmed the species of 349 individuals for the *E. grandis* HS family and 369 individuals for the *E. urophylla* HS family.

2.4.3 SNP genotyping and identification of informative SNPs

Next we performed stringent filtering criteria to identify informative SNP markers. We reclustered the 349 confirmed GU hybrid samples of the *E. grandis* HS family, to obtain an accurate genotypic assignment. For the SNP markers assayed, we obtained a mean GenTrain score of 0.69 and a mean call rate of 0.93. We identified a total of 23 241 polymorphic markers out of the 64 639 SNPs assayed. Of these, 2124 were informative (heterozygous) for the pollen parent (and homozygous in all of the seed parents, Supplementary File 2.1). For the *E. urophylla* HS family, we reclustered SNP data of the 369 confirmed GU samples. We obtained a mean GenTrain score of 0.68 and mean call rate of 0.94. We identified a total of 23 787 polymorphic markers in the mapping dataset. Of these SNPs, 2015 were seed-parent informative (heterozygous in the seed parent and homozygous in all of the pollen parents, Supplementary File 2.1).

2.4.4 Genetic linkage maps

We constructed genetic linkage maps for the *E. grandis* pollen parent and *E. urophylla* seed parent separately. Of the 2124 pollen-parent informative makers and the 2015 seed-parent informative markers, we identified 430 and 362 markers respectively as identical (no recombination) and removed them using JoinMap® 4.1 (Supplementary Files 2.2, 2.3). We defined a total of 11 linkage groups at a logarithm of the odds score (LOD) of 5.0 (*E. grandis*) and LOD of 4.0 (*E. urophylla*) with 1694 and 1653 markers included respectively. The final full genetic linkage maps contained 1610 and 1584 markers for *E. grandis* and *E. urophylla* parents, resulting in a total map length of 896 cM and 982 cM respectively (Table 2.2, 2.3; Supplementary Figure 2.2, 2.3; Supplementary Files 2.4, 2.5). The average marker interval for both of the parental full genetic maps was 0.6 cM. There was high contiguity between the physical and genetic linkage maps for both parents, but there were some markers with marker order changes (Supplementary Figure 2.2, 2.3). We found that a total of 32 markers for the *E. grandis* map and 20 for the *E. urophylla* map, mapped to different linkage groups

than was expected based on the physical positions in the *E. grandis* v2 genome sequence (Supplementary Table 2.3, 2.4; https://phytozome.jgi.doe.gov/).

We constructed framework genetic linkage maps which contained a total of 388 (*E. grandis*) and 422 (*E. urophylla*) markers (Figure 2.1, 2.2; Table 2.2, 2.3; Supplementary files 2.3, 2.4). The total map lengths were 898 cM (*E. grandis*) and 978 cM (*E. urophylla*), resulting in an average marker interval of 2.4 cM for both framework genetic linkage maps. The largest marker interval was 11.1 cM (*E. grandis*) and 7.52 cM (*E. urophylla*). We found that the marker order was conserved between the genetic and physical maps (Supplementary Figure 2.4, 2.5). We identified a total of four markers in the *E. grandis* and three markers for the *E. urophylla* framework genetic linkage maps, which mapped to a different linkage group compared to what was expected based on their physical positions in the *E. grandis* v2 genome sequence (Supplementary Table 2.3, 2.4; https://phytozome.jgi.doe.gov/). These markers could indicate small fragments which are not assembled correctly in the *Eucalyptus* v2 genome or which are duplicated within the genome.

We identified a total of 14.95% and 29.38% of the markers in the framework genetic linkage maps for *E. grandis* and *E. urophylla* HS families respectively, which showed significant deviation from the expected Mendelian segregation ratio at a 0.05 significance level. We expected this as segregation distortion is common in interspecific crosses (Myburg *et al.* 2003; Brondani *et al.* 2006). Additionally, Zuo *et al.* (2019) showed that segregation distortion has little effect on the construction of genetic linkage maps Therefore, we retained markers showing significant segregation distortion in the genetic linkage maps as removing them would have resulted in a large gaps in the maps, especially in regions where there were clusters of distorted markers.

We compared the framework genetic linkage maps with the physical position map to determine the map coverage. Based on the physical positions of the SNP makers, the genetic linkage maps captured

a total of 598.67 Mbp and 599.4 Mbp for *E. grandis* and *E. urophylla* respectively. This resulted in a 99.98% of the genome within a distance of 10 cM from the closest DNA marker.

2.4.5 Trait data

We analysed the following traits; DBH, height, volume, wood density and S:G (trait data summarised in Supplementary Table 2.5, 2.6). We observed a site effect from the distribution of the trait data (Supplementary Figure 2.6, 2.7). The results of the two-way and one-way ANOVA's showed that site and/or family were significantly affecting traits within the *E. grandis* HS and the *E. urophylla* HS families (Supplementary Table 2.7, 2.8, 2.9, 2.10).

We standardised the data for site effect. The standardised trait data is summarised in Table 2.4 and Table 2.5. The two-way and one-way ANOVA's for the *E. grandis* and *E. uropylla* HS families showed that there was no significant effect of site, while family as well as site and family affected some traits (Supplementary Table 2.11, 2.12, 2.13, 2.14).

Next we analysed the standardised data for normality using the Shapiro-Wilk test (Table 2.4, 2.5). We observed that most of the traits were not normally distributed, but there were some traits which were normally distributed on some sites. We also analysed the standardised data correlation using Spearman's and Pearson's correlation (Supplementary Figure 2.8). We found that for both the *E. grandis* HS family and *E. urophylla* HS family, DBH, height and volume were highly correlated with $r > 0.81$ and $rho > 0.75$. This is expected as these traits are dependent on each other.

2.4.6 QTL mapping

We used the standardised phenotypic data for QTL mapping. The permutation test for the significant threshold resulted in a LR threshold of 12.9 for both HS families and was applied across all sites. We analysed QTLs in the two HS families and for each HS family within sites that had more than 80

individual. We included site (for each HS family) and family (for each HS family and each HS family across the different sites) as other traits (o-traits, Supplementary File 2.6). We did not observe any difference in the results when using the cofactors versus not using the cofactors. We mapped the QTLs detected onto the framework genetic linkage maps (Figure 2.1, 2.2).

For the *E. grandis* HS family, we detected QTLs for all traits except for S:G, however, we did not detect QTLs for all of the traits across each site (Figure 2.1, Table 2.6). We observed that the percentage of phenotypic variance explained by the detected QTLs ranged from 5.01% (across the entire HS family) to 36.58% (on site168, Table 2.6). We found that the QTL profiles varied greatly across sites with no QTLs for the same trait overlapping across the sites (Supplementary Figure 2.9). For the *E. urophylla* HS family, we detected QTLs for all traits with the same QTL for DBH, height and volume detected on linkage group 3 across all sites and the whole HS family (Figure 2.2; Table 2.7). Additionally, we found a QTL for wood density which co-located with the QTLs for DBH, height and volume on linkage group 3, across the *E. urophylla* HS family and on site 167. The allelic effect for all these QTLs were in the same direction (Table 2.7; Supplementary Figure 2.10). We did not observe other QTLs for a single trait showing overlap across sites, and the QTL profiles varied greatly across sites (Supplementary Figure 2.10). The percentage of phenotypic variance explained by the QTLs ranged from 3.06% (across the entire FS family) to 19.71% (on site167,Table 2.7). These results show that we could detect medium to large effect QTLs.

We compared the location of detected QTLs with regions of significant segregation distortion across the HS families and the sites. We found that the QTLs for height (Chr10) across the entire *E. grandis* HS family and the QTL for NIR predicted dissolving pulp yield (Chr11) for the *E. grandis* HS family in site165, were in regions of significant distortion. In the *E. uropylla* HS family and across all sites, the QTLs detected on Chr3 for DBH, height and volume were in regions of significant distortion. Additionally, we found that the QTLs for wood density on Chr3 and Chr6 as well as the QTLs for

S:G lignin ratio in the full *E. urophylla* HS family and the QTLs for wood density (Chr3) for site167 were in regions of significant distortion.

Nest we analysed the trait data of individuals and the number of individuals of each genotypic class to determine if a biological reason could explain the growth QTLs identified in regions of significant segregation distortion (Supplementary Figure 2.11). For the *E. grandis* height QTL identified in a significantly distorted region, we found that the genotypic class which was more prevalent in the population actually had smaller trait values, opposite to expectation. For all of the *E. urophylla* growth QTL detected in significantly distorted regions, we found that the genotypic classes with higher trait values were the genotypic classes which were favoured within the population. These results suggest that the hybrid compatibility barriers underlying segregation distortion in the *E. urophylla* FS families could be due to a postzygotic factor which affects early seedling survival and has a growth effect in the surviving individuals.

2.5 Discussion

Genetic linkage maps can be used to identify molecular markers underlying traits of interest which has applications such as MAB. To improve the power and resolution for identifying marker-trait associations, multi-parent mapping populations have been used in many crop species (Buckler *et al.* 2009; Kover *et al.* 2009). Towards this, we constructed framework genetic linkage maps of one *E. grandis* pollen parent and one *E. urophylla* seed parent of a *Eucalyptus* F_1 hybrid breeding trial and mapped QTLs controlling growth and wood properties.

The main limitation of this study was the small sample size at the site and FS family level. The sample sizes are of heightened concern in this study due to the complexity of the population used. The population contained a number of nested FS families and was planted across sites. In order to take into consideration both FS family and site effects, larger sample sizes are needed, with a balanced

representation of FS families per site. We could not perform QTL mapping within all sites as there were not enough individuals present on all the sites. Even within the sites which we did map QTLs, the sample size ranged from 83 to 114 individuals. This limited the power to detect QTLs and the effect of FS family was not taken into consideration.

2.5.1 Identification of informative markers

Informative markers in this study were classified as heterozygous in the common (HS) parent and the same homozygous class across all other parents, which limited the number of informative markers identified. Difficulty in identifying informative markers can be due to the common parent having homozygous markers where heterozygous markers are required. Additionally, markers in the other parents need to be the same homozygous class across all of the parents which is not always possible. This resulted in some regions in the genetic linkage maps having larger marker intervals than desired (> 10 cM). However, despite these limitations, we were still able to identify enough informative markers to have a high average marker density in the full genetic linkage maps and good coverage in the framework linkage maps for QTL detection.

2.5.2 Genetic linkage maps

The full and framework genetic linkage maps for both parents contained 11 linkage groups which correspond to the haploid number of chromosomes in *Eucalyptus* (n = 11). Previous genetic linkage maps for *E. grandis* and *E. urophylla* had map lengths between 822 – 1815 cM and 886 – 1505 cM respectively (Gan *et al.* 2003; Brondani *et al.* 2006; Kullan *et al.* 2012b; Bartholomé *et al.* 2015). The framework genetic linkage maps in this study had total map lengths of 896 cM (*E. grandis*) and 982 cM (*E. urophylla*) which are towards the lower end of map lengths for *Eucalyptus*. This could mean that this study was able to generate more accurate genetic linkage maps as they also have a high contiguity with the physical position maps which suggests accurate marker placements. However, different mapping programmes, algorithms, markers used and mapping populations can influence

map length and could be the reason why there some variation in map length between different studies (Kullan *et al.* 2012b). Additionally, the positions of markers at the end of chromosomes can affect the map length as there may not be sufficient marker coverage at the ends of the chromosomes. In this study, we had the physical positions of the last markers on the genetic linkage maps and the average physical distance from the terminal genetic markers was 0.6 Mbp for both parental maps. This shows that the terminal ends of the chromosomes do have sufficient marker coverage, making the total map length more accurate.

Despite the high contiguity between the genetic and physical positions, there were some markers in both parental maps which mapped to a different linkage group to what was expected based on the *Eucalyptus* v2 genome (Supplementary Table 2.3 and 2.4). This could be due to a number of reasons; the genome sequence assembly containing these markers may be incorrect, the markers may be replicated within the genome, or the markers may be transposed within the parental genomes. Further investigations such as re-sequencing the parental genomes could help determine the cause of the differences.

The comparison of the framework genetic linkage maps with the physical position maps allowed us to evaluate genome coverage. All markers included in the genetic linkage maps have unique map positions within the *Eucalyptus* v2 genome, of which 641 Mb is assembled into contigs. Therefore, 99.98% of the genome was within 10 cM of the nearest marker for both framework genetic linkage maps. This is similar to the genome coverage of Bartholome *et al* 2015 (97.2% - 98.3%) which shows that we were able to identify enough informative markers to achieve a high map coverage.

Segregation distortion occurs when the observed genotypic ratios differ from what is expected under Mendelian segregation ratios. Markers in both the *E. grandis* and *E. urophylla* framework linkage maps showed significant segregation distortion (at a 0.05 significance level). Segregation distortion

in F_1 hybrids can be caused by both pre- and postzygotic hybrid incompatibility factors. In *Eucalyptus*, segregation distortion has previously been reported for interspecific crosses (Grattapaglia and Sederoff 1994; Myburg *et al.* 2003; Kullan *et al.* 2012b) and was therefore expected in this study. Segregation distortion and hybrid compatibility is discussed in more detail in Chapter 3.

2.5.3 QTL mapping

We were able to detect QTLs for growth and wood properties segregating in the F_1 hybrid progeny, with the percentage of phenotypic variance explained by the QTLs ranging from 3.06 % - 36.5%, which shows that both large effect and moderate effect QTLs were detected. Due to different mapping population and different markers being used in this study it is challenging to compare results with previous studies. There were some linkage groups (LG) which had QTLs for the same traits detected in previous studies, such as LG1 for wood density in *E. grandis* and LG3, LG6, LG8 and LG9 for wood density in *E. urophylla* (Kullan *et al.* 2012a). However, we cannot determine if these are the same QTLs due to the different marker systems used in the studies.

Due to the population being generated as an F_1 hybrid breeding trial, we were initially only going to perform QTL mapping in the entire HS families which had a sufficient number of individuals to have a high statistical power. However, the discovery of genotype-by-environment interaction led us to perform QTL mapping within each site. When mapping within the sites, the power was limited due to the small sample size across each site, which varied between 83 – 141 individuals. This can be seen by QTL peaks which do not reach the significance threshold (Supplementary Figure 2.9, 2.10). While the percentage of variation explained by QTL detected in this study ranged from 3.06% – 36.5%, the smallest percentage was obtained for the *E. urophylla* HS family across all sites which had the largest number of individuals (n = 367). With the small population size, the Beavis effect also applies which means that the percentage of variation explained by a QTL may be overestimated (Beavis 1998). Therefore, we expect that an increase in sample size in each site will allow for the

detection of more small effect QTLs and greater precision for the allele effects that are detected. Wood density had the largest number of QTLs detected in both the *E. grandis* (8 QTLs) and *E. urophylla* HS (9 QTLs) as well as within the sites. These results are consistent with previous studies which found a higher number of QTLs for wood density than DBH (Kullan *et al.* 2012a). This is due to growth traits such as height, volume and DBH having a lower heritability than wood density, which results in smaller effect QTLs (Kullan *et al.* 2012a).

Previous studies found that segregation distortion can have an effect on QTL mapping. Xu (2008) found that segregation distortion had a negative effect on QTL detection 44% of the time for QTLs with dominant effects. Zhang *et al.* (2010) found that the power of QTL detection can be reduced in significantly distorted regions. However, the same study concluded that in large populations, the effect on the position and false-positive detection rate of QTLs would not be affected by segregation distortion. In the current study, some of the QTLs detected were found to be in significantly distorted regions. Due to the small sample size of this study, there is a chance that the position of the QTLs detected in those regions are false-positives. Future studies would need to be conducted in larger populations to confirm these results.

We hypothesised that segregation distortion could have an effect on growth traits. To determine if there could be a biological reason for the QTLs detected in significantly distorted regions, we analysed the trait data and number of individuals for each genotypic class for the detected growth QTLs in distorted regions. For the *E. urophylla* QTLs, the genotypic classes that were more prevalent contained individuals with higher growth trait values. This could mean that the hybrid incompatibility factor (or hybrid viability factor) causing the segregation distortion is providing a growth advantage, or more likely the individuals carrying the alternative allele are have their growth affected and are therefore smaller. For the *E. grandis* height QTL on chromosome 10, the opposite was seen. Here individuals carrying the more prevalent allele were smaller than the individuals carrying the

alternative allele. This QTL should therefore be treated with caution as it could be a false positive QTL caused by the segregation distortion. These results suggest that including significantly distorted loci in QTL mapping studies can help to better understand the effect of hybrid incompatibility on growth traits.

2.5.4 Genotype-by-environment effect on QTL

GxE has been detected in previous studies in *Eucalyptus* when analysing QTLs across different environments. The GxE interaction has been identified for both growth and wood properties (e Silva *et al.* 2006; Freeman *et al.* 2013). In the current study, different QTLs were detected across individual sites which suggests that GxE plays a role. For the *E. grandis* HS family, there was no overlap between QTLs identified on the different sites. This shows that there could be environment dependent interactions which affect growth and wood property traits. We however, could not directly estimate the GxE effects on QTLs and sample size limitations made QTL detection within single sites prone to false-negatives.

QTLs for DBH, height and volume were detected on chromosome 3, for the *E. urophylla* HS family and across the two sites. These traits were highly correlated (r = 0.81 – 0.96, which is expected as they are mathematically dependent on each other) so it can be expected that the QTLs would be located on the same region of the genome. The detection of QTLs in the same region across the different sites suggests that these QTLs are expressed on all sites and possibly not affected by the environment. Therefore, markers underlying this region could be possible targets for marker-assisted breeding. Additionally, we found a QTL for wood density which co-located with the QTLs for DBH, height and volume on chromosome 3 for the *E. urophylla* HS family and across site 167. These QTLs all had allelic effects in the same direction. Wood density was also found to be positively correlated with DBH, height and volume (r = 0.30 – 0.36). These results suggest that higher growth is associated with higher wood density. Therefore, within this QTL region, there could be one gene controlling all

of the above traits or multiple tightly linked genes. Previous studies have also identified the co-location of QTLs for growth traits and wood density in *Eucalyptus* (Grattapaglia *et al.* 1996, Thumma *et al.* 2010). However, Thumma *et al.* (2010) found that there was a negative correlation between wood density and DBH. Therefore, there could be many complex interactions between multiple genes which affect growth and wood density traits. The QTLs for all the wood properties in the *E. urophylla* HS family were site specific which suggests that these traits are affected by environment.

While we were able to map QTLs across sites, the effect of FS family was not taken into consideration. During QTL mapping we did add FS family as a co-factor but this did not change the outcome of the analysis. This could be due to the small sample sizes on each site and limited representation of FS families per site. Therefore, future studies will need to be done using balanced FS families (equal number of progeny per FS family) so that the effect of family can be taken into account.

2.6 Conclusions

We were able to construct framework genetic linkage maps in a *Eucalyptus* multi-parent population with an average marker interval of 2.4 cM for an *E. grandis* pollen parent and *E. urophylla* seed parent. QTLs were identified for growth and wood properties in HS families and within different sites, with each QTL explaining between 3.06 % and 36.58%,. This study shows that a multi-parent mapping approach can work in an outcrossing species such as *Eucalyptus*. However, larger sample sizes are required at the FS family and site level, with balanced FS families. The ability to use this approach in *Eucalyptus* is advantageous as these types of populations already exist as F_1 hybrid breeding trials and can now be exploited for genetic dissection of quantitative traits.

2.7 Acknowledgements

This work was funded by Sappi Forest research (Hilton, KZN, South Africa), the Department of Science and Technology (DST), Technology Innovation Fund (TIA) and the National Research Foundation (NRF). JC acknowledges the NRF for MSc bursary support.

2.8 Tables

Table 2.1 *Eucalyptus* multi-parent mapping population. Nine *E. grandis* pollen parents were crossed with eight *E. urophylla* seed parents. This study focused on one *E. grandis* half-sib (HS) family and one *E. urophylla* HS family (bold).

			E. grandis pollen parents									
			FK592	FK604	FK593	FK605	FK594	**FK595**	FK596	FK597	FK608	
		Provenances[a]	Australia[b]	Australia[b]	Australia[b]	SAFRI[c]	SAFRI[c]	**SAFRI[c]**	Zululand[d]	Zululand[d]	Zululand[d]	*E. urophylla* HS individuals
E. urophylla seed parents	FK606	Lomblen	0	0	0	60	0	**112**	0	112	24	308
	FK598	Lomblen	0	0	102	0	63	**75**	0	103	0	343
	FK607	Lomblen	53	0	0	0	30	**0**	74	22	0	179
	FK599	Timor	115	111	105	79	104	**52**	113	109	0	788
	FK600	Timor	77	48	26	0	105	**111**	82	75	0	524
	FK601	Flores	132	113	104	47	51	**113**	59	30	0	649
	FK602	**Flores**	**84**	**60**	**84**	**108**	**107**	**110**	**81**	**0**	**0**	**634**
	FK603	Flores	103	103	97	25	30	**81**	30	48	0	517
		E. grandis HS individuals	564	435	518	319	490	**654**	439	499	24	

[a] Provenances show where the original seed was sourced from. The *E. urophylla* seed parents had seed sourced from Indonesian Islands as shown
[b] Seed sourced from a seed orchard in Australia
[c] Seed sourced from the South African Forestry Research Institute (SAFRI) *E. grandis* breeding program
[e] Seed sourced from *E. grandis* bred specifically for the Zululand regions by the South African pulp and paper industry (Sappi) breeding program

Table 2.2 Summary of SNP markers mapped for each linkage group in the *E. grandis* HS family

Linkage group	*E. grandis* HS family									
	Full genetic linkage map						Framework genetic linkage map			
	No. of informative markers	No. of informative markers after identical markers removed	No. markers mapped	Size in cM	Mean distance between markers (cM)	Largest marker interval (cM)	No. markers	Size in cM	Mean distance between markers (cM)	Largest marker interval (cM)
1	181	144	143	77.54	0.54	3.33	35	76.95	2.26	3.86
2	251	175	168	83.13	0.50	10.49	31	82.67	2.76	10.44
3	120	94	90	89.02	1.00	11.77	27	92.18	3.55	11.06
4	176	136	111	61.09	0.56	8.90	29	63.43	2.27	4.88
5	210	172	165	74.74	0.46	5.08	35	74.73	2.20	5.68
6	273	227	227	102.86	0.46	2.71	52	102.71	2.01	3.61
7	167	149	145	61.28	0.43	5.01	29	61.34	2.19	5.04
8	237	194	191	109.56	0.68	5.59	49	107.87	2.25	5.54
9	146	121	98	68.68	0.71	9.16	25	68.17	2.84	8.74
10	158	127	124	78.74	0.64	5.47	35	78.47	2.31	5.54
11	205	155	148	89.08	0.61	4.66	41	89.24	2.23	4.63
Total	2124	1694	1610	895.71	0.60	11.77	388	897.75	2.44	11.06

Table 2.3 Summary of SNP markers mapped for each linkage group in the *E. urophylla* HS family

Linkage group	*E. urophylla* HS family									
	Full genetic linkage map						Framework genetic linkage map			
	No. of informative markers	No. of informative markers after identical markers removed	No. markers mapped	Size in cM	Mean distance between markers (cM)	Largest marker interval (cM)	No. markers	Size in cM	Mean distance between markers (cM)	Largest marker interval (cM)
1	195	152	149	80.89	0.55	4.73	36	80.68	2.31	4.74
2	139	120	119	81.61	0.69	6.74	34	81.45	2.47	6.90
3	198	158	155	104.79	0.68	7.21	40	104.16	2.67	7.52
4	154	129	126	73.53	0.59	5.43	33	73.68	2.30	5.48
5	100	92	90	72.06	0.81	4.32	31	71.95	2.40	4.31
6	280	225	198	116.68	0.59	4.30	52	115.87	2.27	5.44
7	137	123	117	75.12	0.65	3.83	35	74.76	2.2	4.36
8	247	200	192	109.39	0.57	6.76	42	108.73	2.65	6.80
9	213	172	166	75.90	0.46	3.18	35	75.59	2.22	3.70
10	175	120	137	90.42	0.67	4.76	40	89.75	2.30	4.53
11	177	138	135	101.29	0.76	4.43	44	100.90	2.35	5.60
Total	2015	1629	1584	981.68	0.64	7.21	422	977.51	2.38	7.52

Table 2.4 Summary statistics of the corrected trait data of the *E. grandis* HS family. Traits analysed were diameter at breast height (DBH), height, volume, wood density, NIR dissolving pulp yield (dPY) and NIR S:G lignin ratio (S:G). Minimum (min) and maximum values were calculated for each trait in the *E. grandis* HS family and across the four sites. Shapiro-Wilk test for normality was performed for each trait.

E. grandis HS family corrected data (n = 349)				
				Shapiro-Wilk W Test
Trait	Min	Max	W	p-value
DBH	-2.69	3.45	0.97	6.44E-06*
Height	-2.01	4.12	0.89	2.51E-13*
Volume	-4.14	2.1	0.99	8.76E-03*
Density	-2.57	2.91	1	4.96E-02*
dPY	-2.3	5.09	0.92	9.44E-11*
SG	-3.1	3.03	0.99	3.10E-02*
E. grandis Site165 corrected data (n=105)				
DBH	-2.69	2.58	0.96	4.60E-03*
Height	-1.53	3.53	0.87	4.57E-07*
Volume	-4.14	1.67	0.94	9.40E-04*
Density	-2.57	1.74	0.97	7.62E-2
dPY	-2.11	5.09	0.9	6.43E-06*
SG	-2.96	2.43	0.99	8.63E-01
E. grandis Site166 corrected data (n = 62)				
DBH	-1.86	2	0.94	2.12E-02*
Height	-2.1	1.81	0.96	1.06E-01
Volume	-2.19	1.62	0.97	2.24E-01
Density	-1.82	2.37	0.96	8.82E-02
dPY	-1.53	3.15	0.92	3.15E-03*
SG	-3.1	1.89	0.93	6.83E-03*
E. grandis Site167 corrected data (n = 99)				
DBH	-1.85	2.78	0.96	1.15E-02*
Height	-1.52	3.63	0.9	9.84E-06*
Volume	-2.04	2.06	0.98	1.91E-01
Density	-2.32	2.91	0.98	1.70E-01
dPY	-2.3	2.33	0.95	6.04E-03*
SG	-1.88	3.03	0.96	2.85E-02*
E. grandis Site168 corrected data				
DBH	-1.67	3.45	0.94	1.67E-03*
Height	-0.84	4.12	0.73	5.67E-10*
Volume	-2.17	2.1	0.99	7.76E-01
Density	-2.16	2.15	0.97	1.60E-01
dPY	-1.17	3.31	0.79	2.79E-08*
SG	-2.79	1.21	0.84	5.47E-07*

* Statistically significant at P < 0.05

Table 2.5 Summary statistics of the corrected trait data of the *E. urophylla* HS family. Traits analysed were diameter at breast height (DBH), height, volume, wood density, NIR dissolving pulp yield (dPY) and NIR S:G lignin ratio (S:G). Minimum (min) and maximum values were calculated for each trait in the *E. urophylla* HS family and across the four sites. Shapiro-Wilk test for normality was performed for each trait.

E. urophylla HS family corrected data (n = 367)				
			Shapiro-Wilk W test	
Trait	Min	Max	W	p-value
DBH	-2.14	4.56	0.97	1.52E-06*
Height	-1.98	7.32	0.88	2.00E-15*
Volume	-3.24	2.21	0.99	5.98E-03*
Density	-2.95	3.21	1	9.57E-01
dPY	-2.19	4.16	0.96	1.64E-07*
SG	-4.13	3.34	0.99	2.93E-03*
E. urophylla Site165 corrected data (n = 100)				
DBH	-2.14	2.66	0.94	4.60E-04*
Height	-1.31	3	0.85	1.83E-08*
Volume	-3.24	1.66	0.97	2.35E-02*
Density	-2.95	2	0.97	3.99E-02*
dPY	-2.19	3.09	0.99	4.31E-01
SG	-3.82	1.73	0.97	1.83E-02*
E. urophylla Site166 corrected data (n = 67)				
DBH	-1.85	2.49	0.97	2.28E-01
Height	-1.94	2.42	0.98	3.74E-01
Volume	-2.2	2.76	0.97	1.46E-01
Density	-1.98	1.6	0.96	6.43E-01
dPY	-1.67	4.16	0.91	9.13E-04*
SG	-4.13	1.91	0.9002	5.58E-04*
E. urophylla Site167 corrected data (n = 141)				
DBH	-1.94	4.56	0.96	5.35E-04*
Height	-1.98	7.32	0.86	9.02E-10*
Volume	-2.36	1.99	0.98	1.50E-02*
Density	-2.26	3.21	0.99	4.74E-01
dPY	-1.58	3.19	0.92	2.08E-06*
SG	-1.96	3.34	0.95	1.86E-04*
E. urophylla Site168 corrected data (n = 59)				
DBH	-1.54	3.48	0.93	4.789E-03*
Height	-0.76	4.46	0.65	1.04E-09*
Volume	-2.03	2.21	0.98	7.45E-01
Density	-2.16	2.15	0.98	4.35E-01
dPY	-1.36	2.84	0.85	4.61E-05*
SG	-2.4	1.99	0.96	1.41E-01

* Statistically significant at $P < 0.05$

Table 2.6 QTL detected for the *E. grandis* HS family pollen parent. QTL analysis was performed across three sites and the entire HS family. QTL were detected (at a 0.05 threshold) for height, wood density, diameter at breast height (DBH), volume, and NIR dissolving pulp yield (dPY). Composite interval mapping (CIM) in QTLCartographer v1.17 (Basten et al 1998, 2000) was used to detect QTL.

Trait	Linkage group	Peak position (cM)	LR	Additive effect	Variance (R^2) explained by QTL (%)
E. grandis HS (n = 349)					
Height	10	42.59	15.63	-0.45	5.01
Wood density	1	3.11	16.67	0.45	5.10
Wood density	10	21.80	24.29	-0.78	7.50
E. grandis site165 (n = 105)					
DBH	8	105.04	12.97	0.66	10.44
Wood density	9	58.23	17.75	0.79	14.32
dPY	5	51.13	13.74	0.69	11.02
dPY	11	19.02	18.95	1.21	13.87
E. grandis site167 (n = 99)					
DBH	11	37.87	14.99	1.19	14.47
Volume	3	29.85	15.13	0.99	15.86
Wood density	1	3.11	16.53	0.74	13.26
Wood density	6	52.73	26.74	-1.00	23.05
E. grandis site168 (n = 83)					
Volume	3	40.92	32.88	-1.55	36.58
Wood density	8	38.73	22.61	1.45	21.51
Wood density	10	21.80	21.62	-1.10	22.44
Wood density	10	50.91	17.90	0.95	17.44

Table 2.7 QTL detected for the *E. urophylla* HS family seed parent. QTL analysis was performed across two sites and across the HS family. QTL were detected (at a 0.05 threshold) for diameter at breast height (DBH), height, volume, wood density, NIR dissolving pulp yield (dPY) and NIR S:G lignin ratio (S:G). Composite interval mapping (CIM) in QTLCartographer v1.17 (Basten et al 1998, 2000) was used to detect QTL.

Trait	Linkage group	Peak position (cM)	LR	Additive effect	Variance (R^2) explained by QTL (%)
E. urophylla HS (n = 367)					
DBH	2	70.05	12.92	0.48	3.06
DBH	3	53.70	70.52	-0.94	18.25
Height	3	53.70	40.50	-0.76	10.71
Volume	3	53.70	66.34	-0.90	17.05
Wood density	1	2.33	15.19	0.42	4.36
Wood density	3	42.02	14.97	-0.43	4.10
Wood density	3	64.48	17.26	-0.46	4.71
Wood density	6	60.47	13.59	0.40	3.78
Wood density	11	70.69	14.04	-0.41	4.06
dPY	11	94.70	24.61	-0.53	6.88
S:G	4	6.65	16.12	-0.63	4.93
S:G	6	50.46	17.36	-0.48	5.01
E. urophylla site165 (n = 100)					
DBH	3	53.70	27.35	-0.90	19.30
Height	3	53.70	23.75	-1.22	18.49
Volume	3	53.70	22.89	-0.81	15.50
Wood density	8	89.12	14.63	-0.65	11.05
Wood density	11	75.84	25.18	-0.86	19.68
E. urophylla site167 (n = 141)					
DBH	3	49.54	34.22	-0.98	18.96
Height	3	51.54	20.29	-0.84	11.37
Volume	3	49.54	37.28	-0.98	19.71
Wood density	3	63.97	16.52	-0.62	9.02
Wood density	9	32.57	19.62	-1.07	10.52
dPY	11	79.30	13.38	-0.57	8.02

2.9 Figures

Figure 2.1 (Legend on page 63)

Figure 2.1 Framework genetic linkage map with QTL segregating in the *E. grandis* HS family pollen parent. The maps were constructed in JoinMap® 4.1 (Van Ooijen 2005) and visualized using LinkageMapView 2.1.2 (Ouellette et al. 2018). A total of 388 markers were mapped over 11 linkage groups, resulting in a total map length of 898 cM. The average marker distance is 2.4 cM and the largest marker interval is 11 cM. The marker positions are shown on the left of each linkage group (cM Kosambi) and marker names are shown on the right of each linkage group. QTLs were identified for the following traits; diameter at breast height (DBH), height, volume, wood density, NIR dissolving pulp yield (dPY) and NIR S:G lignin ratio (S:G). QTLCartographer v1.17 (Basten et al 1994, 2004) was used to identify significant QTL (genome-wide threshold of 0.05), using composite interval mapping (CIM). The colour of the bars represent the test population in which the QTL were detected; green is the entire *E. grandis* HS family, pink is site165, purple is site167 and orange is site168.

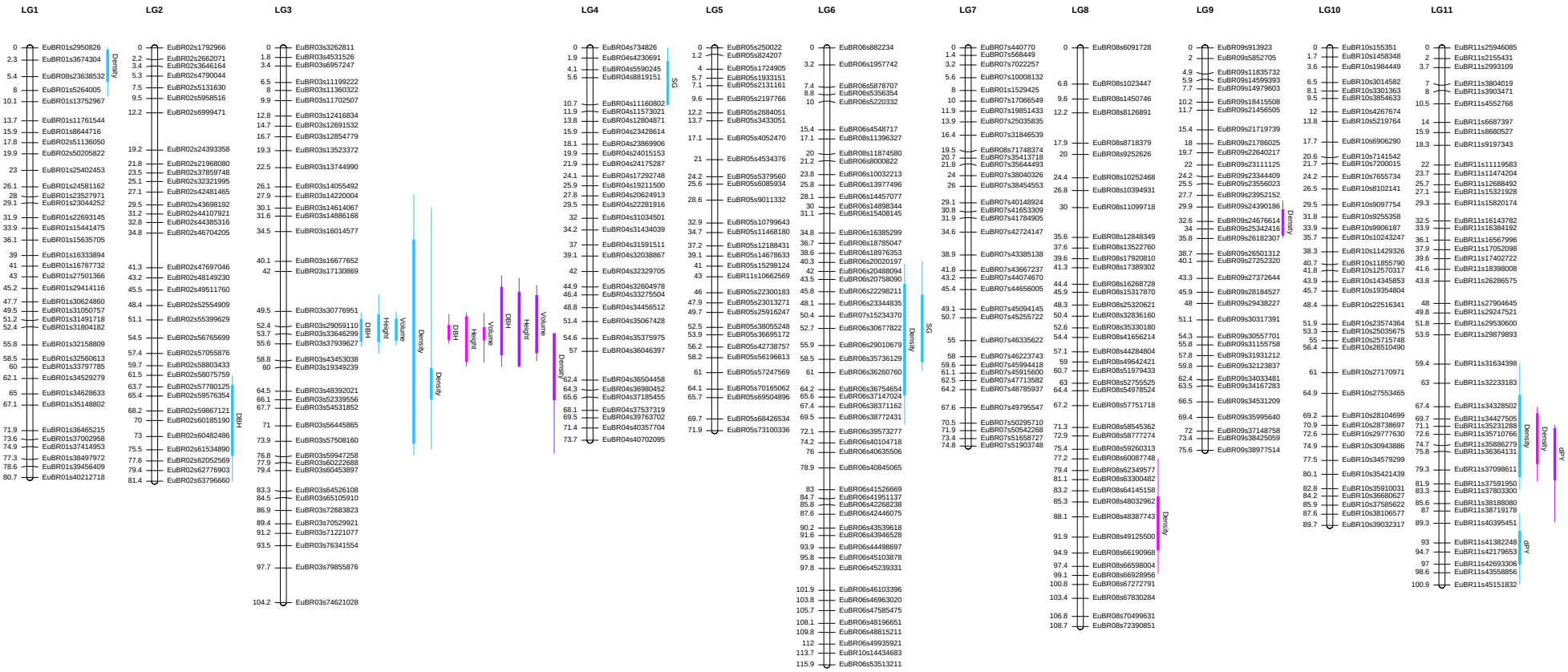


Figure 2.2 (Legend on page 63)

Figure 2.2 Framework genetic linkage map with QTL segregating in the *E. urophylla* HS family. The maps were constructed in JoinMap®4.1 (Van Ooijen 2006) and visualised using LinkageMapView 2.1.2 (Ouellette et al. 2018). A total of 11 linkage groups, containing 422 markers were mapped. Marker names are shown on the left and marker positions on the right (cM, Kosambi). The largest marker interval was 7.5 cM and the average marker distance was 2.4 cM. QTL were identified for the following traits; diameter at breast height (DBH), height, volume, wood density, NIR dissolving pulp yield (dPY) and NIR S:G lignin ratio (S:G) in the *E. urophylla* HS family seed parent. QTLCartographer v1.17 (Basten et al 1994, 2004) was used to identify significant QTL (genome-wide threshold of 0.05), using composite interval mapping (CIM). The colour of the bars represent the population in which the QTL were detected; blue is the entire *E. urophylla* HS family, pink is site165 and purple is site167.

2.10 References

de Assis TF. 2000. Production and use of *Eucalyptus* hybrids for industrial purposes. In: Dungey H, Dieters M, Nikles D, eds. QFRI/CRC-SPF Symposium. Noosa, Queensland, Australia: Department of primary industries, 63–74.

Bartholomé J, Mandrou E, Mabiala A, Jenkins J, Nabihoudine I, Klopp C, Schmutz J, Plomion C, Gion J. 2015. High-resolution genetic maps of *Eucalyptus* improve *Eucalyptus grandis* genome assembly. *New Phytologist* 206: 1283–1296.

Basten CJ, Weir BS, Zeng Z-B. 1994. Zmap-a QTL cartographer. In: Smith C, Gavora JS, Benkel B, Chesnais J, Farifull W, Gibson JP, Kennedy BW, Burnside EB, eds. Proceedings of the 5th World Congress on Genetics Applied to Livestock Production: Computing Strategies and Software. Guelph, Ontario, Canada: Organizing committee, 5th World Congress on Genetics Applied to Livestock Production, 65–66.

Basten CJ, Weir BS, Zeng Z-B. 2004. QTL Cartographer. North Carolina State University.

Beavis W. 1998. QTL analyses: power, precision, and accuracy. In: Paterson A, ed. Molecular dissection of complex traits. Boca Raton FL, 145–162.

Broman KW, Gatti DM, Simecek P, Furlotte NA, Prins P, Sen Ś, Yandell BS, Churchill GA. 2019. R/qtl2: Software for mapping quantitative trait loci with high-dimensional data and multiparent populations. *Genetics* 211: 495–502.

Brondani RP V, Williams ER, Brondani C, Grattapaglia D. 2006. A microsatellite-based consensus linkage map for species of *Eucalyptus* and a novel set of 230 microsatellite markers for the genus. *BMC Plant Biology* 6: 20.

Buckler ES, Holland JB, Bradbury PJ, Acharya CB, Brown PJ, Browne C, Ersoz E, Flint-Garcia S, Garcia A, Glaubitz JC, *et al.* 2009. The genetic architecture of maize flowering time. *Science* 325: 714–718.

Costa e Silva JC, Potts BM, Dutkowski GW. 2006. Genotype by environment interaction for growth of *Eucalyptus globulus* in Australia. *Tree Genetics and Genomes* 2: 61–75.

Freeman JS, Potts BM, Downes GM, Pilbeam D, Thavamanikumar S, Freeman J. 2013. Stability of quantitative trait loci for growth and wood properties across multiple pedigrees and environments in *Eucalyptus globulus*. *New Phytologist* 198: 1121–1134.

Gan S, Shi J, Li M, Wu K, Wu J, Bai J. 2003. Moderate-density molecular maps of *Eucalyptus urophylla* S . T . Blake and *E . tereticornis* Smith genomes based on RAPD markers. *Genetica* 118: 59–67.

Grattapaglia D, Sederoff R. 1994. Genetic linkage maps of *Eucalyptus grandis* and *Eucalyptus*

urophylla using a pseudo-testcross: Mapping strategy and RAPD markers. *Genetics* 137: 1121–1137.

Grattapaglia D, Bertolucci FLG, Penchel R, Sederoff RR. 1996. Genetic mapping of quantitative trait loci controlling growht and wood quality traits in *Eucalyptus grandis* using a maternal half-sib family and RAPD markers. *Genetics* 144: 1205-1214

Grattapaglia D, Silva-Junior OB, Resende RT, Cappa EP, Müller BSF, Tan B, Isik F, Ratcliffe B, El-Kassaby YA. 2018. Quantitative genetics and genomics converge to accelerate forest tree breeding. *Frontiers in Plant Science* 871: 1–10.

Hirschhorn JN, Daly MJ. 2005. Genome-wide association studies for common diseases and complex traits. *Nature Reviews Genetics* 6: 95–108.

Huang BE, George AW. 2011. R/mpMap : a computational platform for the genetic analysis of multiparent recombinant inbred lines. *Bioinformatics* 27: 727–729.

Jones OR, Wang J. 2010. COLONY: A program for parentage and sibship inference from multilocus genotype data. *Molecular Ecology Resources* 10: 551–555.

Kalinowski ST, Taper ML, Marshall TC. 2007. Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. *Molecular Ecology* 16: 1099–1106.

Kover PX, Valdar W, Trakalo J, Scarcelli N, Ehrenreich IM, Purugganan MD, Durrant C, Mott R. 2009. A multiparent advanced generation inter-cross to fine-map quantitative traits in *Arabidopsis thaliana*. *PLoS Genetics* 5: e1000551.

Kullan ARK, van Dyk MM, Hefer CA, Jones N, Kanzler A, Myburg AA. 2012a. Genetic dissection of growth, wood basic density and gene expression in interspecific backcrosses of *Eucalyptus grandis* and *E. urophylla*. *BMC Genetics* 13: 60.

Kullan ARK, van Dyk MM, Jones N, Kanzler A, Bayley A, Myburg AA. 2012b. High-density genetic linkage maps with over 2,400 sequence-anchored DArT markers for genetic dissection in an F2 pseudo-backcross of *Eucalyptus grandis* × *E. urophylla*. *Tree Genetics and Genomes* 8: 163–175.

Lange K, Boehnke M. 1982. How many polymorphic genes will it take to span the human genomic? *American Journal of Human Genetics* 34: 842–845.

Mackay TF. 2001. The genetic architecture of quantitative traits. *Annual Review of Genetics* 35: 303–339.

Myburg AA, Grattapaglia D, Tuskan GA, Hellsten U, Hayes RD, Grimwood J, Jenkins J, Lindquist E, Tice H, Bauer D, *et al.* 2014. The genome of *Eucalyptus grandis*. *Nature* 510: 356–362.

Myburg AA, Griffin AR, Sederoff R, Whetten R. 2003. Comparative genetic linkage maps of *Eucalyptus grandis*, *Eucalyptus globulus* and their F1 hybrid based on a double pseudo-backcross

mapping approach. *Theoretical and Applied Genetics* 107: 1028–1042.

Thumma BR, Baltunis BS, Bell JC, Emebiri LC, Moran GF, Southerton SG. 2010. Quantitative trait locus (QTL) analysis of growth and vegetative progagation traits in *Eucalyptus nitens* full-sib families. *Tree Geentics & Genomes* 6: 877-889

Van Ooijen JW. 2006. JoinMap ® 4, Software for the calculation of genetic linkage maps in experimental populations.

Ouellette LA, Reid RW, Blanchard SG, Brouwer CR. 2018. LinkageMapView—rendering high-resolution linkage and QTL maps (O Stegle, Ed.). *Bioinformatics* 34: 306–307.

Peng-yuan L, Jun Z, Lu Y. 2006. Impacts of QTL x environment interactions on genetic response to marker-assisted selection. *Acta Genetica Sinica* 33: 63–71.

Retief ECL, Stanger TK. 2009. Genetic parameters of pure and hybrid populations of *Eucalyptus grandis* and *E. urophylla* and implications for hybrid breeding strategy. *Southern Forests: a Journal of Forest Science* 71: 133–140.

Silva-Junior OB, Faria DA, Grattapaglia D. 2015. A flexible multi-species genome-wide 60K SNP chip developed from pooled resequencing of 240 *Eucalyptus* tree genomes across 12 species. *New Phytologist* 206: 1527–1540.

Wingfield MJ, Swart WJ, Abear BJ. 1989. First record of *Cryphonectria* canker of *Eucalyptus* in South Africa. *Phytophylactica* 21: 311–313.

Xu S. 2008. Quantitative trait locus mapping can benefit from segregation distortion. *Genetics* 180: 2201–2208.

Yu J, Holland JB, McMullen MD, Buckler ES. 2008. Genetic design and statistical power of nested association mapping in maize. *Genetics* 178: 539–551.

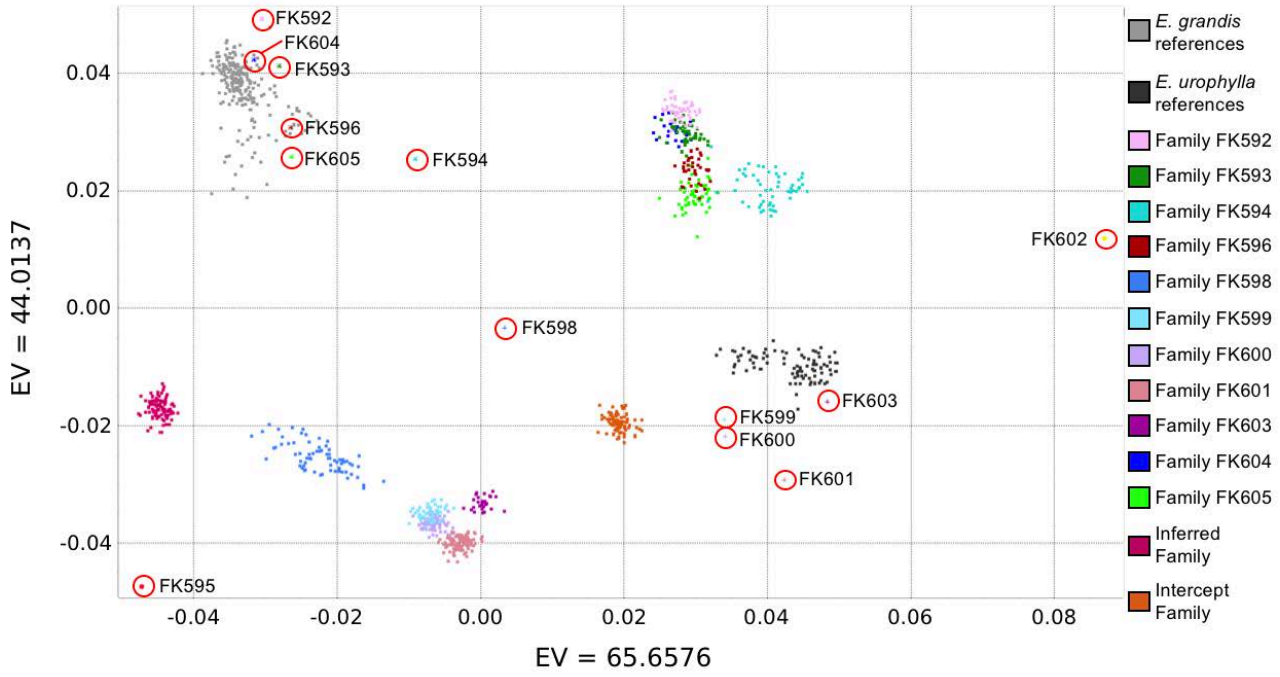
Zeng Z-B. 1993. Theoretical basis for separation of multiple linked gene effects in mapping quantitative trait loci. *Proceedings of the Natural Academy of Sciences USA* 90: 10972–10976.

Zeng Z-B. 1994. Precision mapping of quantitative trait loci. *Genetics* 136: 1457–1468.

Zhang L, Wang S, Li H, Deng Q, Zheng A, Li S, Li P, Li Z, Wang J. 2010. Effects of missing marker and segregation distortion on QTL mapping in F2 populations. *Theoretical and Applied Genetics* 121: 1071–1082.

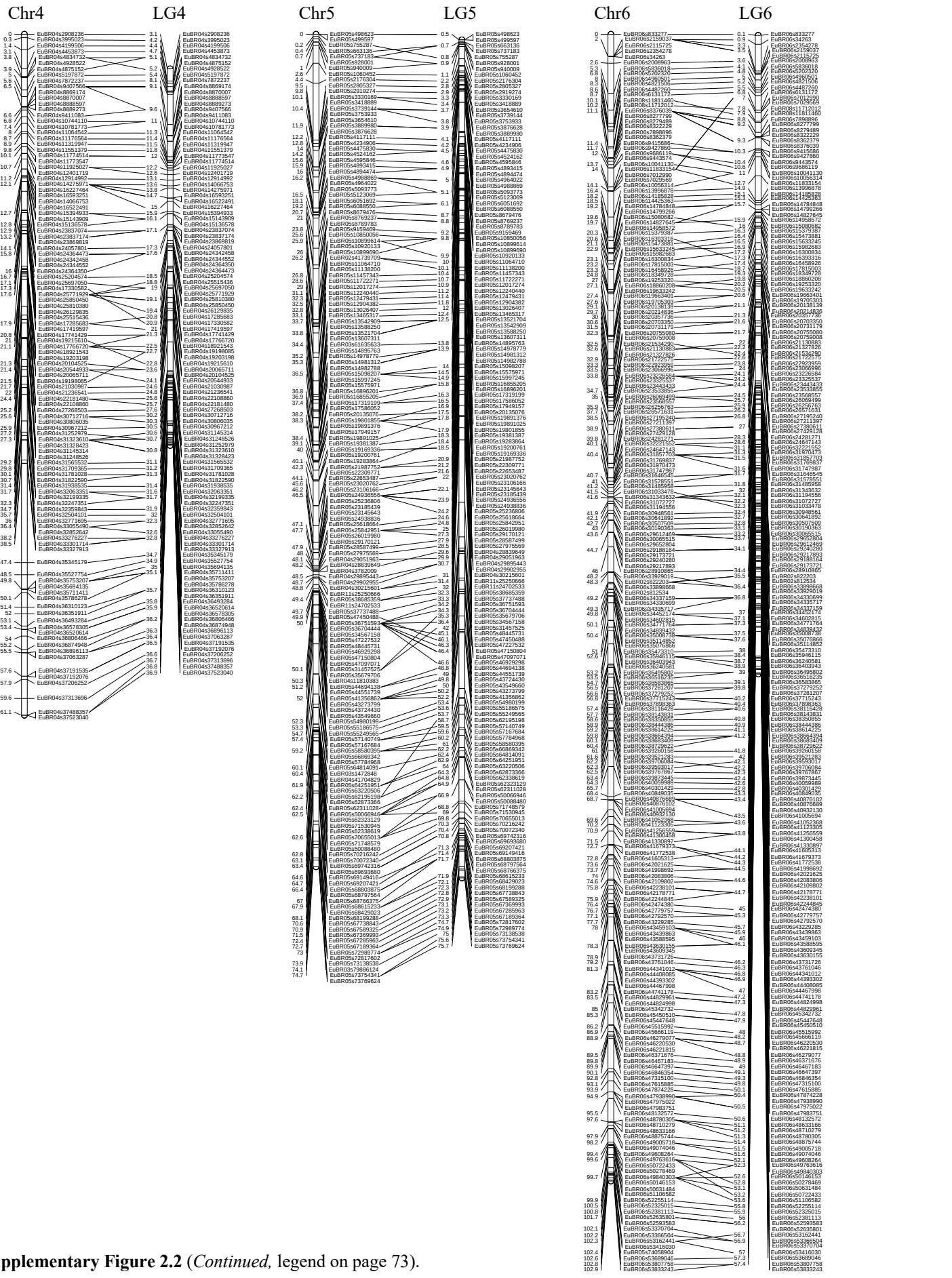
Zheng C, Boer MP, Eeuwijk FA Van. 2019. Construction of genetic linkage maps in multiparental populations. *Genetics* 212: 1031–1044.

2.11 Supplementary Figures

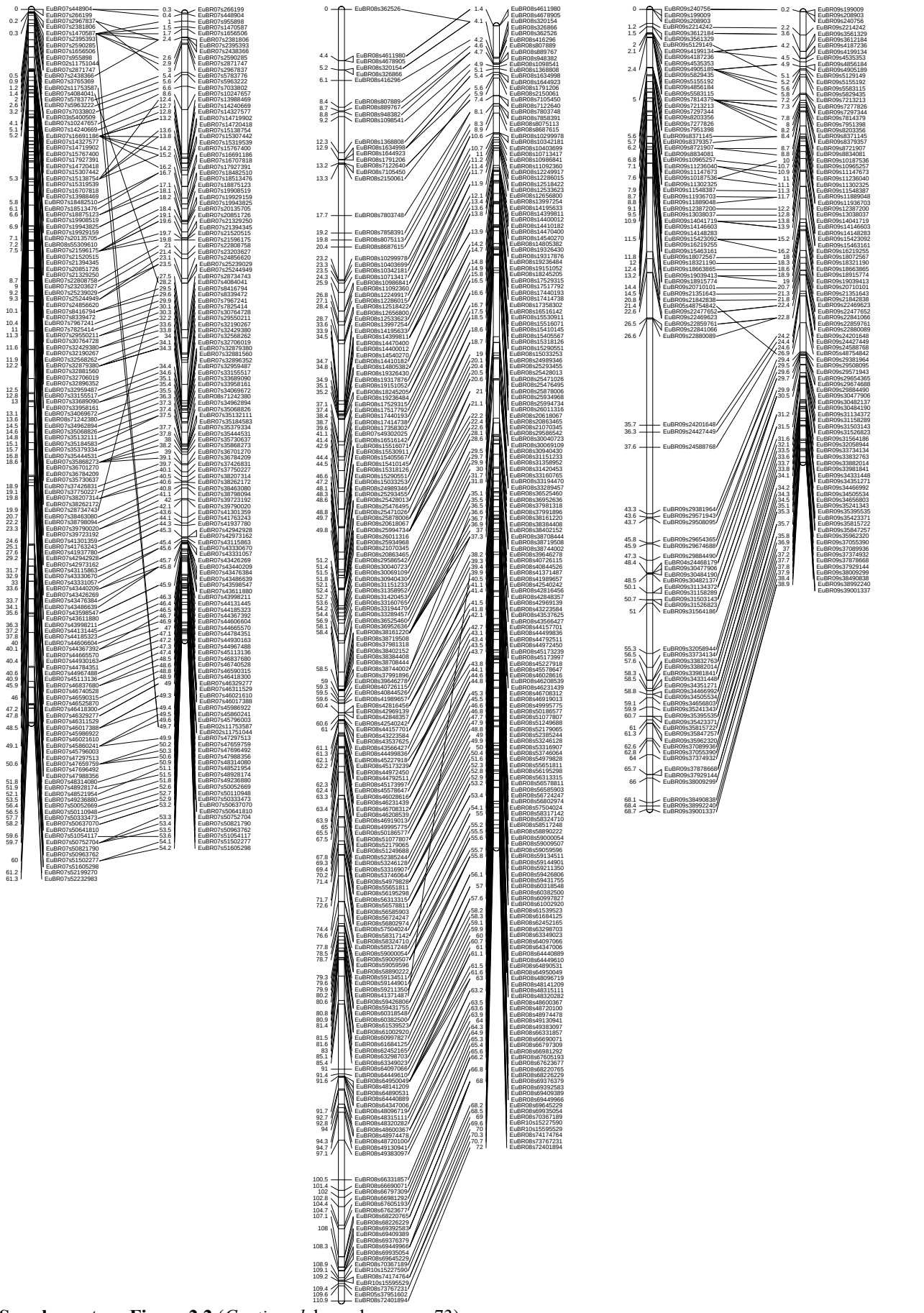


Supplementary Figure 2.1 PCA showing the clustering of SNP genotypes of FS families in the *E. grandis* and *E. urophylla* HS families. *E. grandis* and *E. urophylla* reference sets were used. A total of six FS families in the *E. urophylla* HS family cluster half-way between the *E. urophylla* seed parent (FK602) and the *E. grandis* pollen parents (FK592, FK593, FK594, FK596, FK604, FK605) which suggests that these are *E. grandis* x *E. urophylla* F_1 hybrids. The intercept FS family (orange) cluster half-way between the *E. urophylla* seed parent (FK602) and the *E. grandis* pollen parent (FK595) which is consistent with the family being a GU F_1 hybrid. Of the seven *E. grandis* FS families, four clustered between the *E. grandis* pollen parent (FK595) and the *E. urophylla* seed parents (FK599, FK600, FK601, and FK603) consistent with being GU F_1 hybrids. One *E. urophylla* seed parent (FK598) is half-way between the *E. grandis* and *E. urophylla* references which suggests it is a GU F_1 hybrid. The FS family FK598 clusters half-way between seed parent FK598 and the *E. grandis* pollen parent (FK595) which suggests that the FS family is most likely GUxG backcross. The family in which the seed parent was unknown (inferred family) clusters near the *E. grandis* pollen parent (FK595) which suggests the seed parent is a pure *E. grandis* individual and therefore the FS family is the result of a GxG cross. The inferred family and family FK598 were removed from the study, resulting in five FS families within the *E. grandis* HS family.

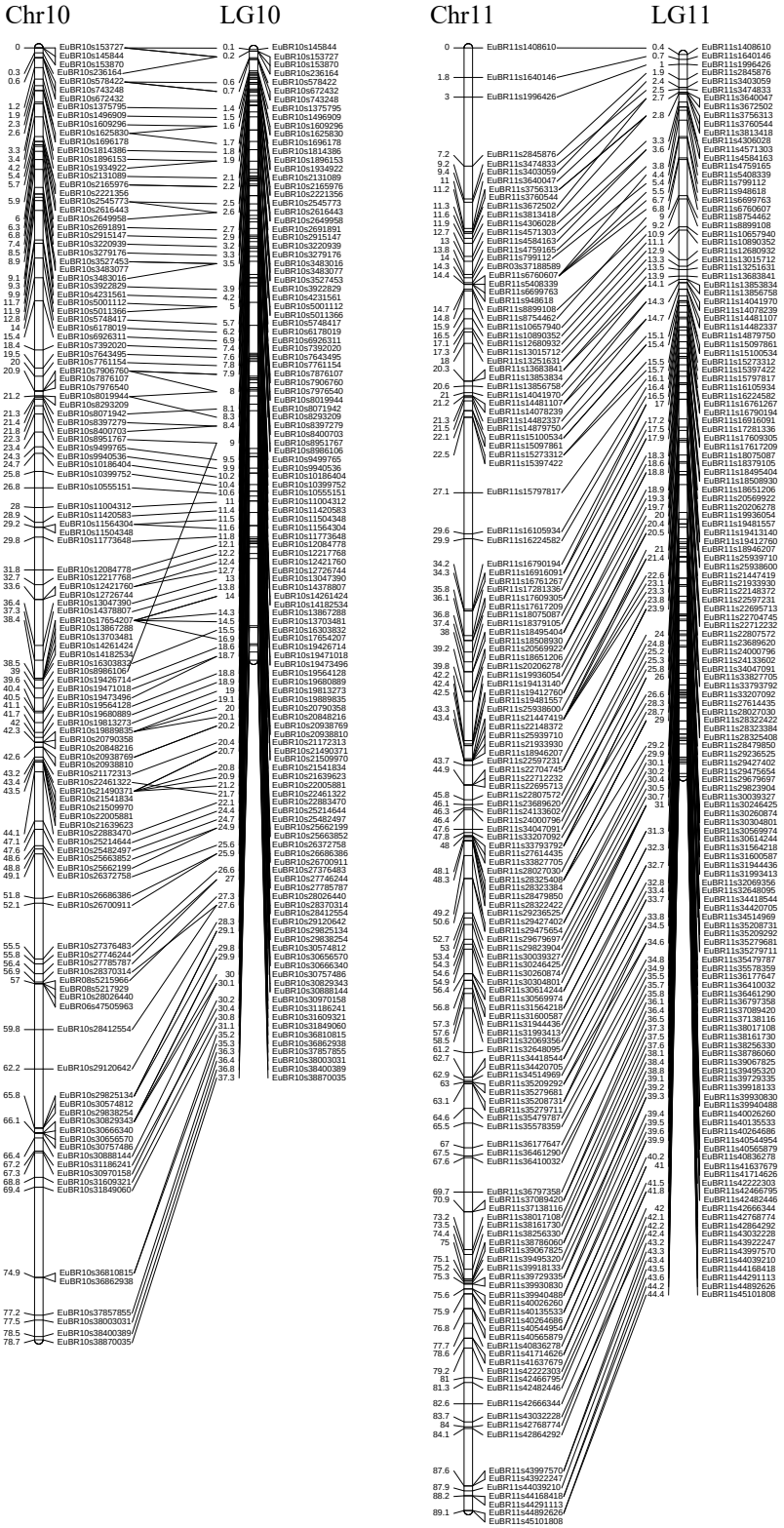
Supplementary Figure 2.2 (Legend on page 73).



Supplementary Figure 2.2 (*Continued,* legend on page 73).

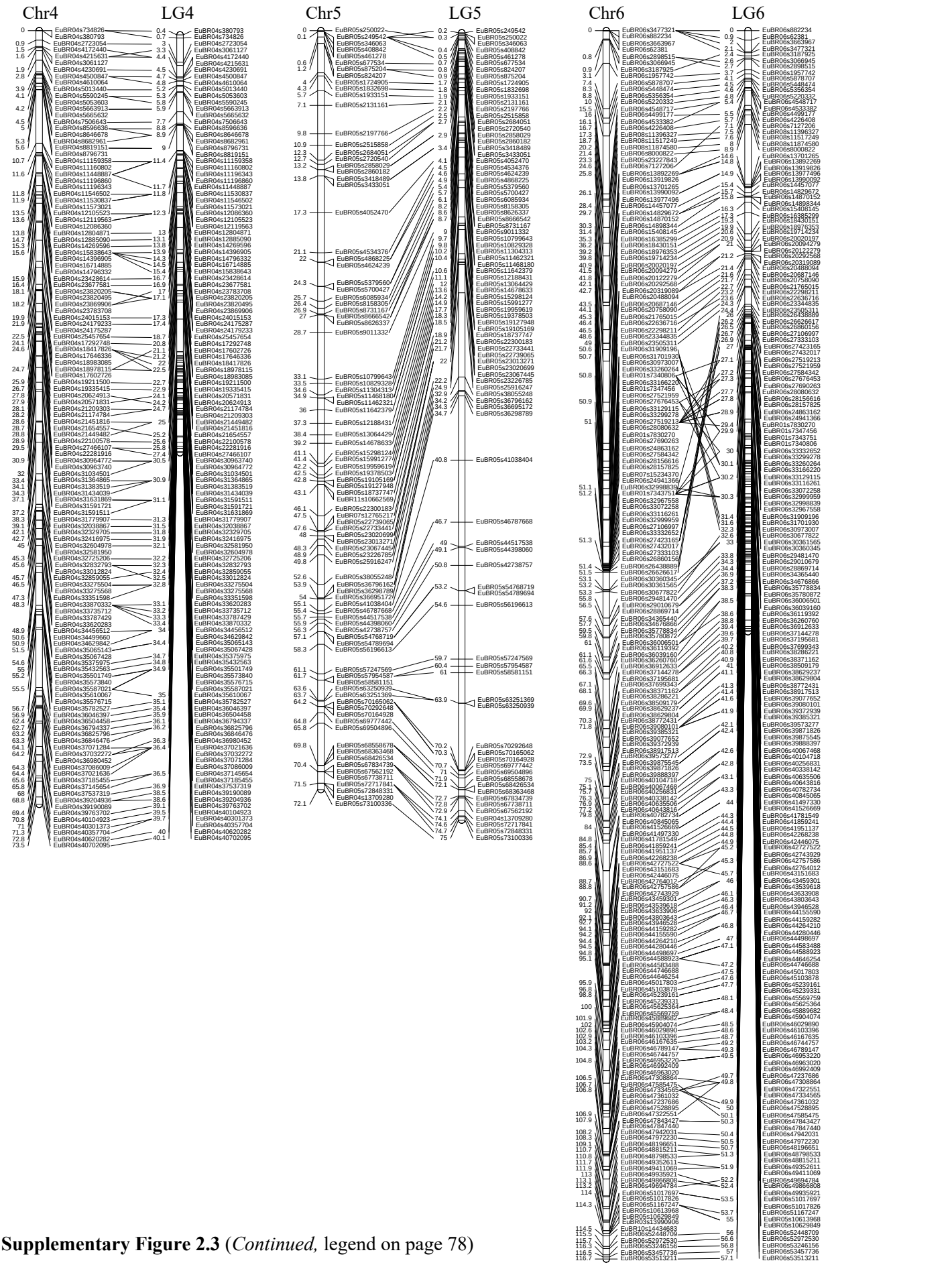


Supplementary Figure 2.2 (*Continued,* legend on page 73).

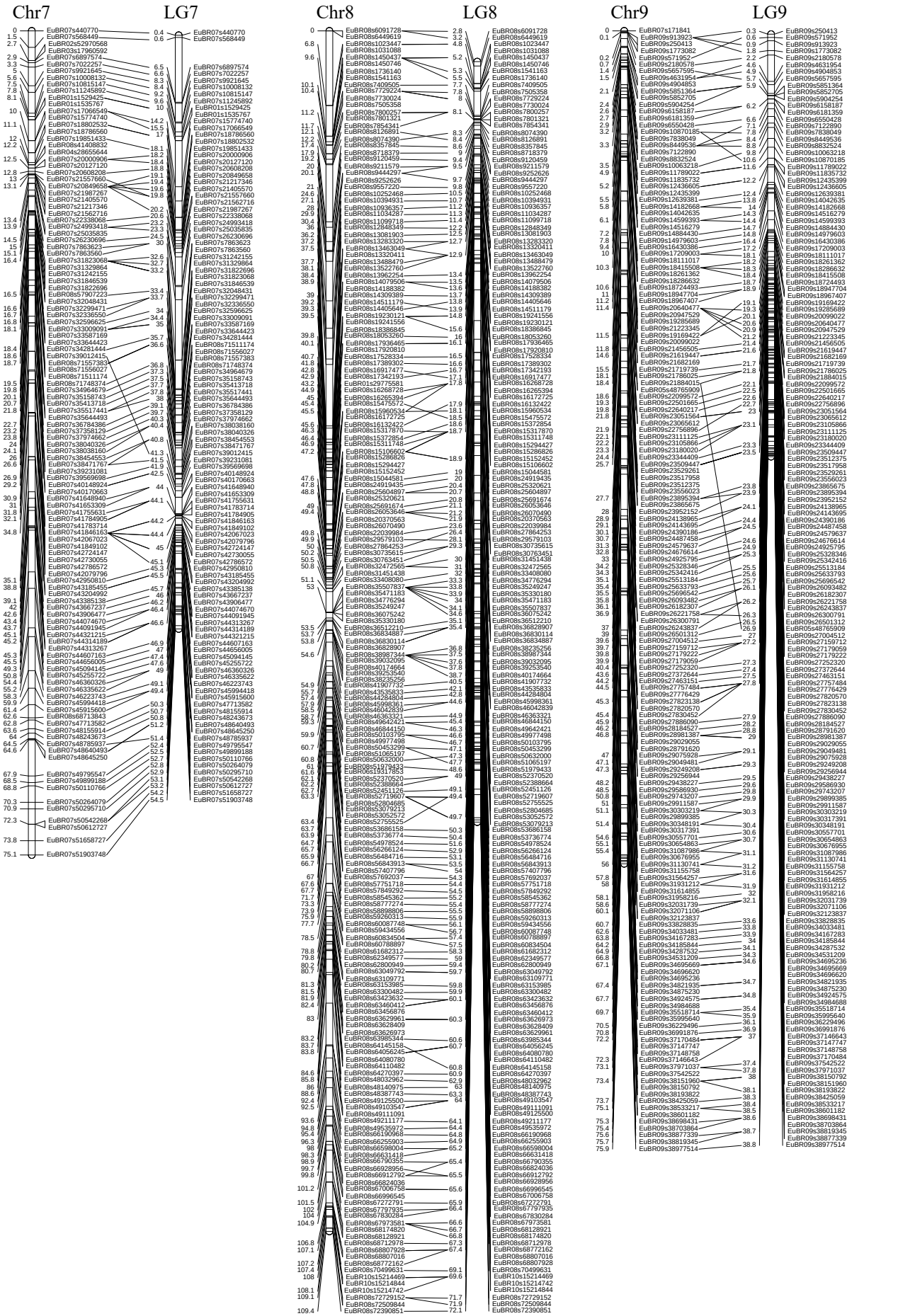


Supplementary Figure 2.2 *E. grandis* full genetic linkage map and physical map. The full genetic linkage map (left) and the physical position map (right) were visualised using LinkageMapView 2.1.2 (Ouellette et al. 2018). The marker positions (cM Kosambi) and marker names are shown on the left and right of the linkage groups respectively. A total of 1610 SNP markers are included in the genetic map resulting in a total map length of 896 cM. The largest marker interval was 11.8 cM and the average.

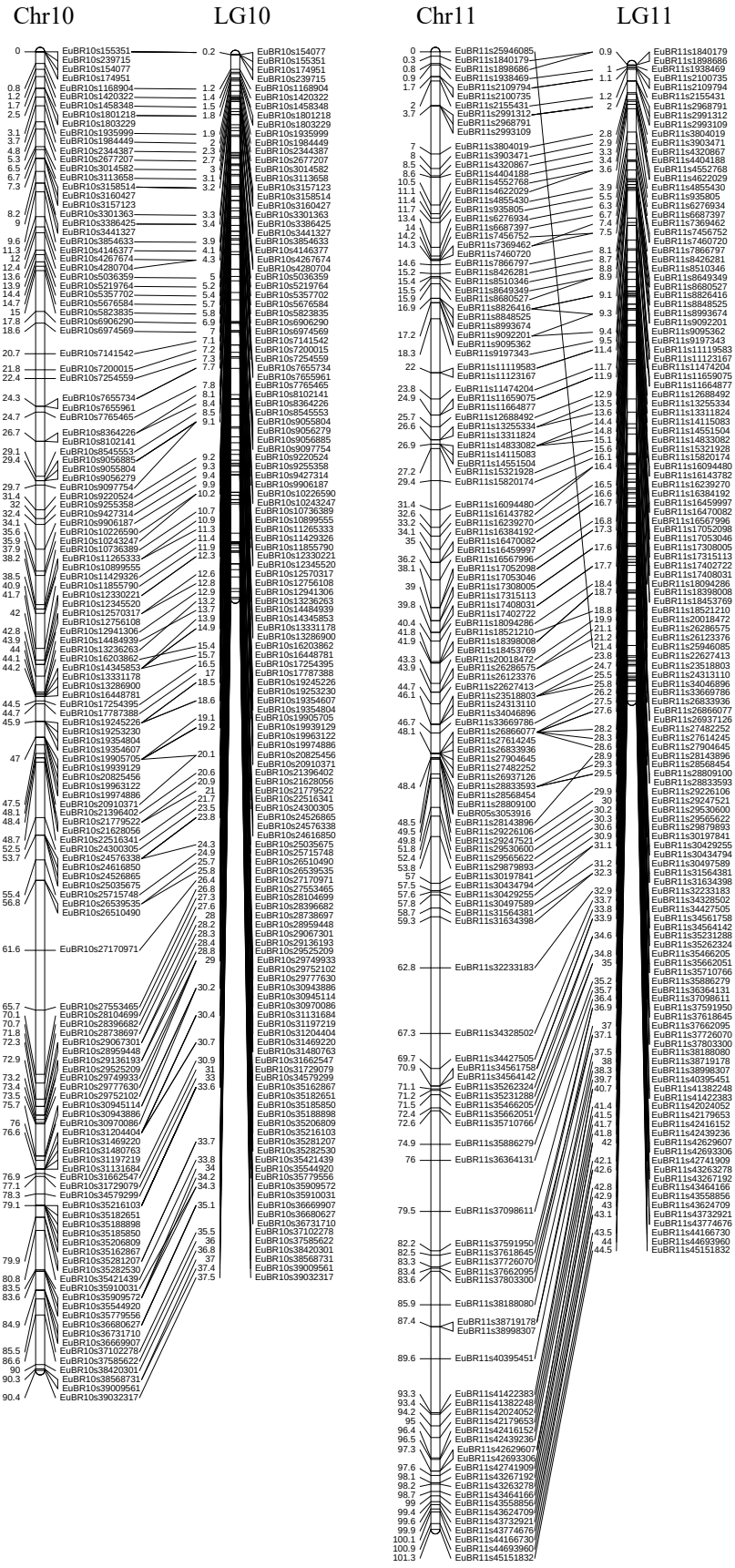
Supplementary Figure 2.3 (Legend on page 78).



Supplementary Figure 2.3 (*Continued,* legend on page 78)

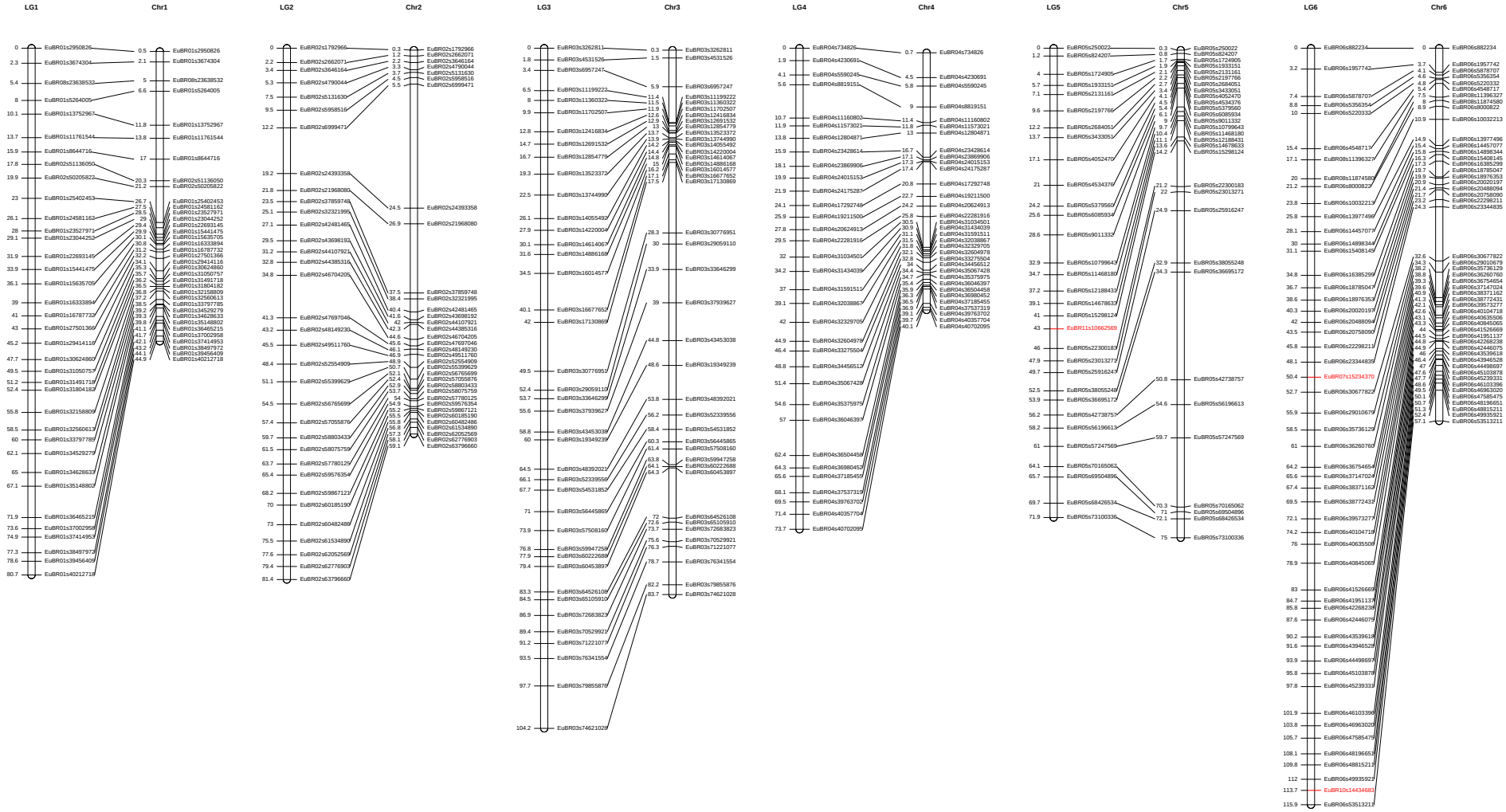


Supplementary Figure 2.3 (*Continued,* legend on page 78*).*

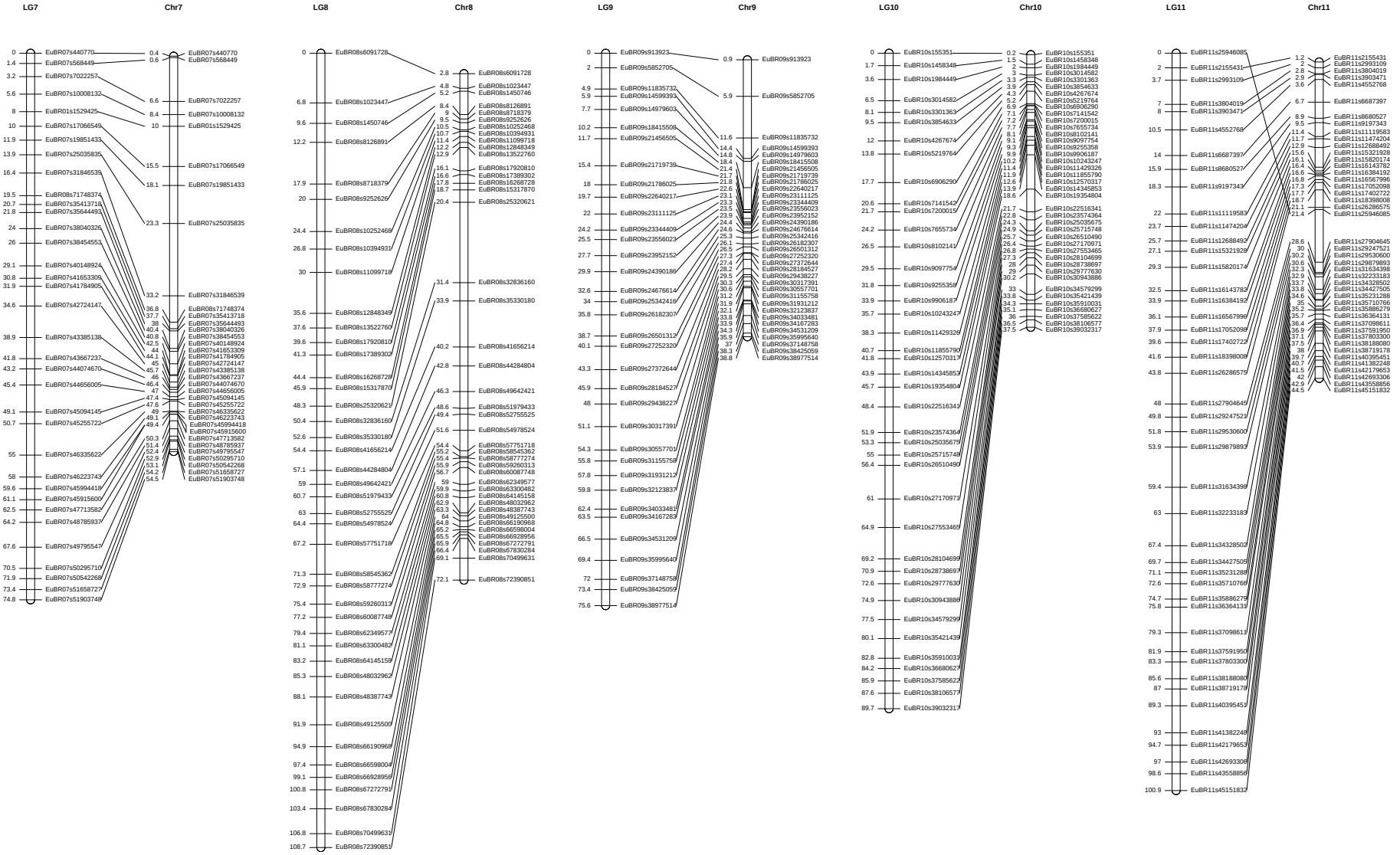


Supplementary Figure 2.3 *(Continued,* legend on page 78*).*

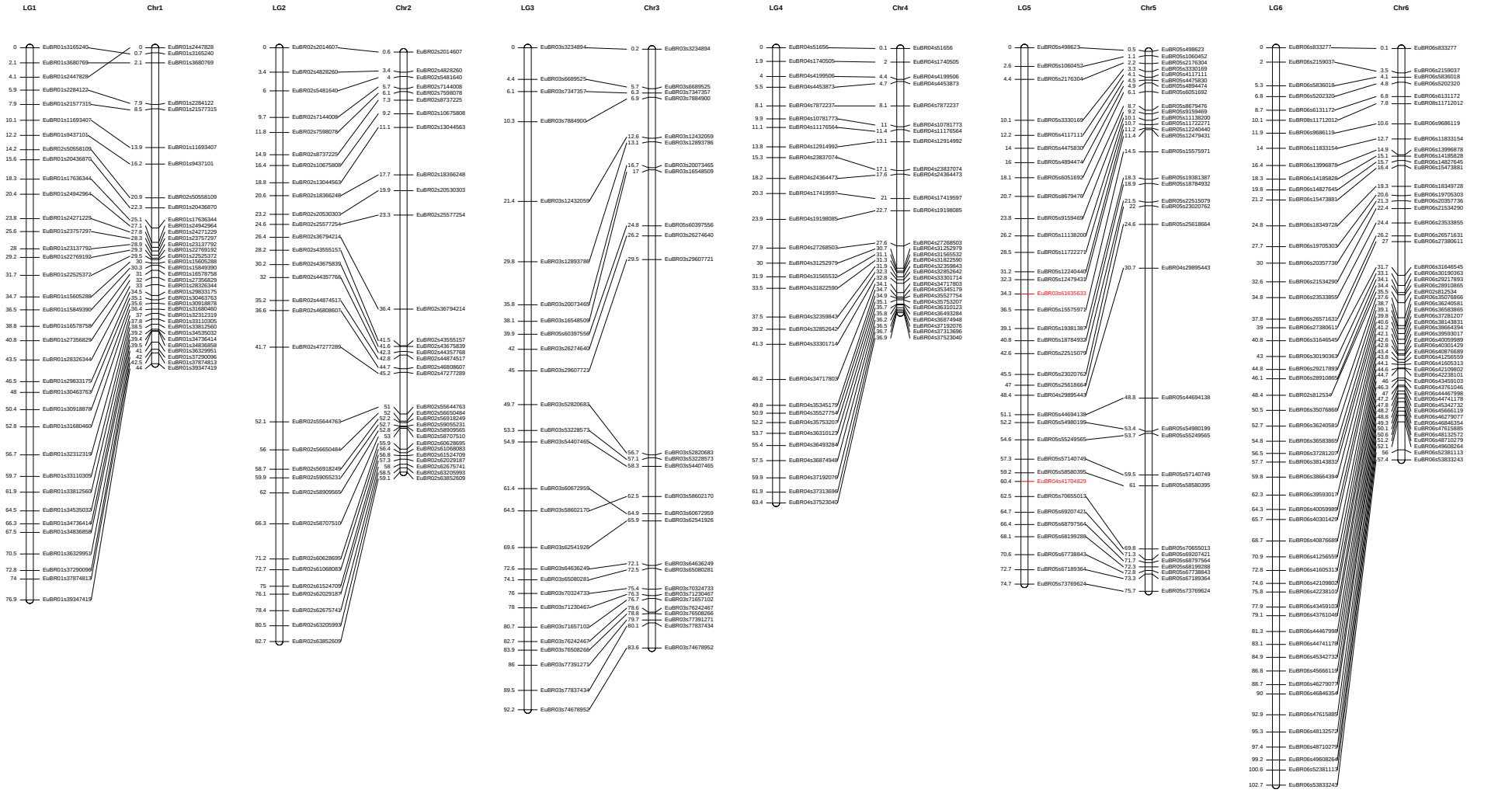
Supplementary Figure 2.3 *E. urophylla* full genetic linkage map and physical map. The full genetic linkage map (right) and the physical position map (left) was constructed using 1653 seed parent informative markers. The marker names are shown on the right with the position (Kosambi, cM) on the left. The genetic linkage maps were visualised using LinkageMapView 2.1.2 (Ouellette et al. 2018). The maps contained 11 linkage groups with a total map distance 982 cM . The largest marker interval was 7.2 cM and the average marker distance was 0.6 cM.



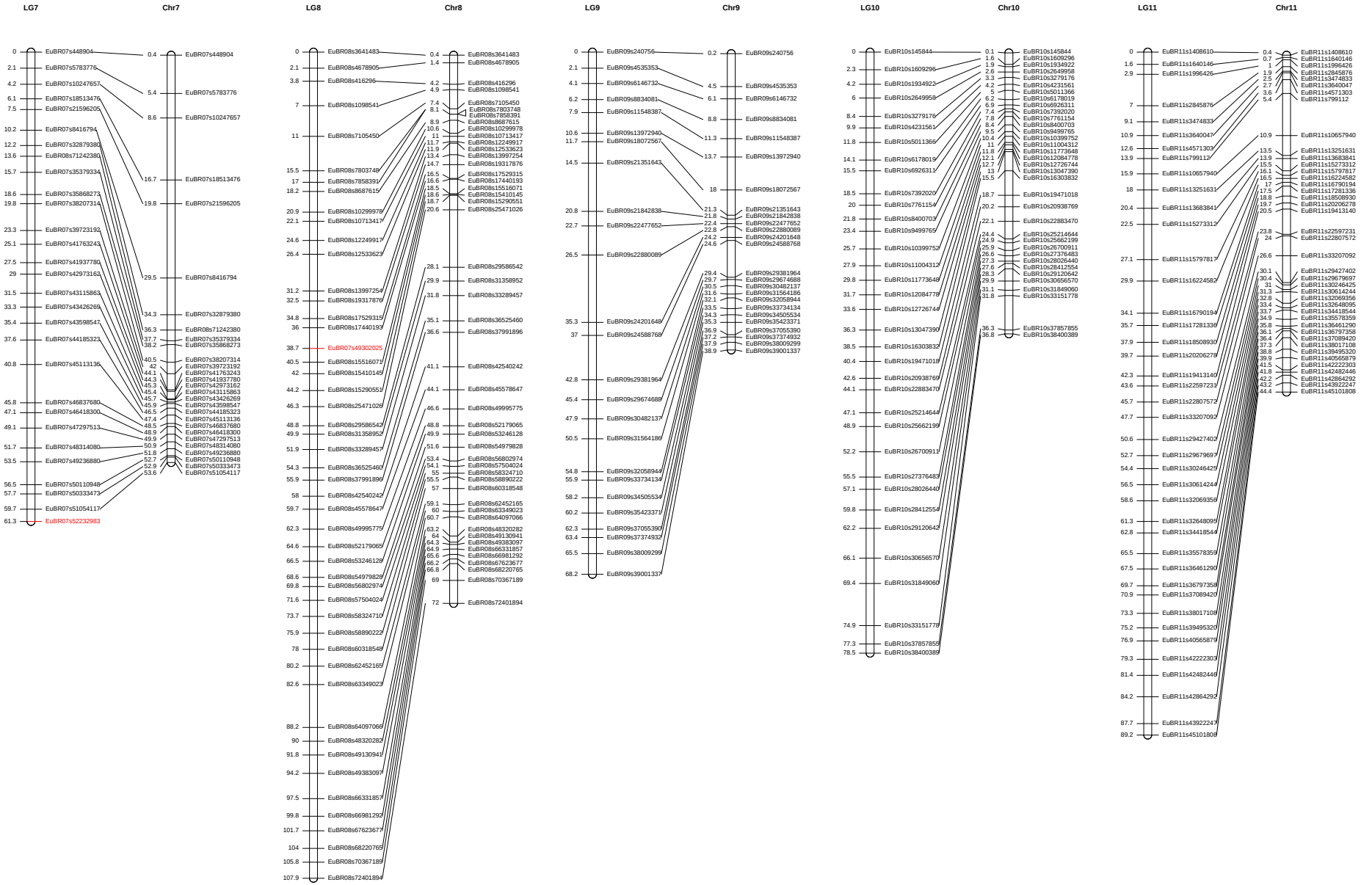
Supplementary Figure 2.4 Framework genetic linkage map and physical map for the *E. urophylla* seed parent. The marker order is generally conserved between the physical map (Mbp, right) and the framework genetic linkage map (cM, left).



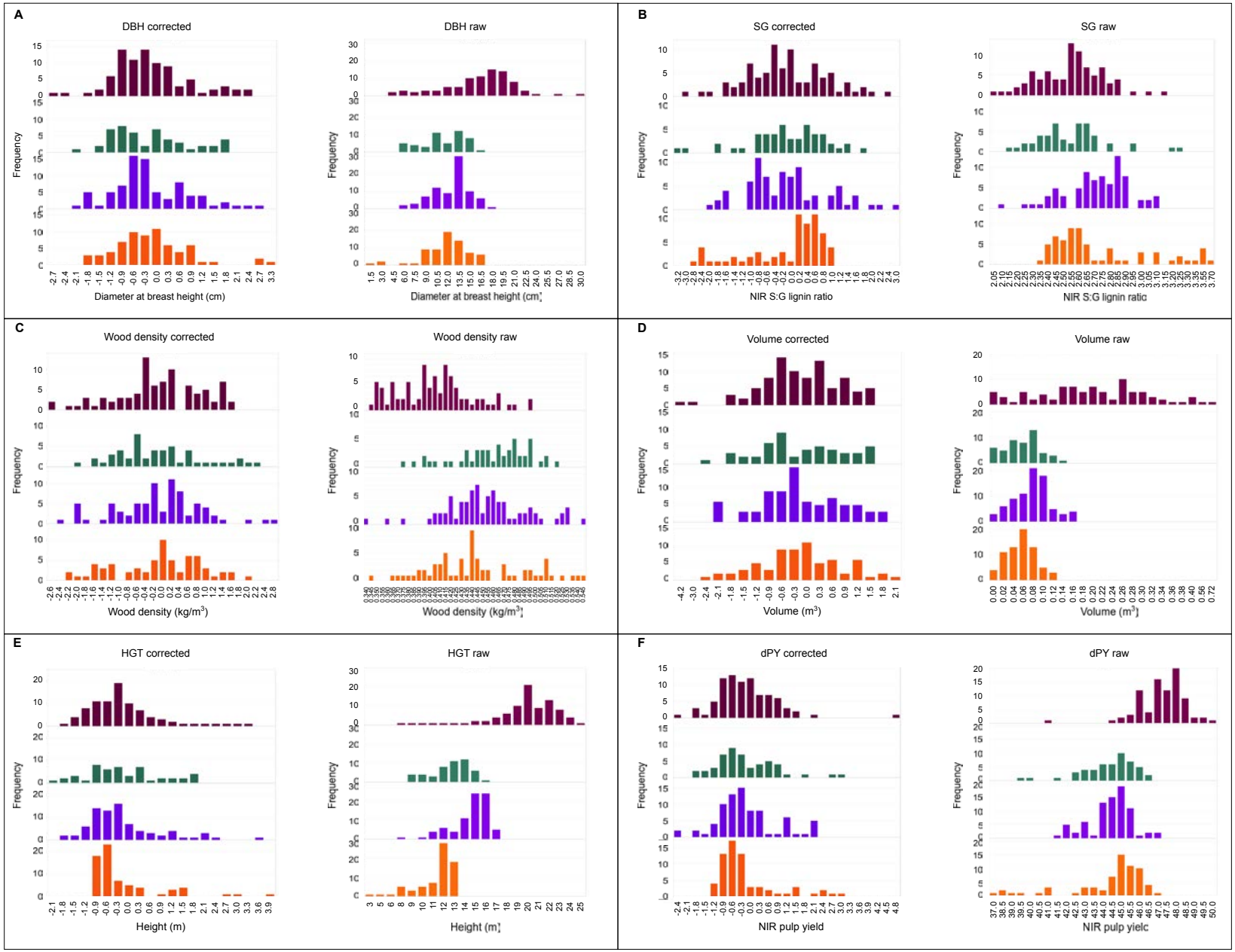
Supplementary Figure 2.4 (*Continued*).



Supplementary Figure 2.5 Framework genetic linkage map and physical position map for *E. grandis* pollen parent. The marker order is generally conserved between the physical position map (Mbp, right) and the framework genetic linkage map (cM, left).

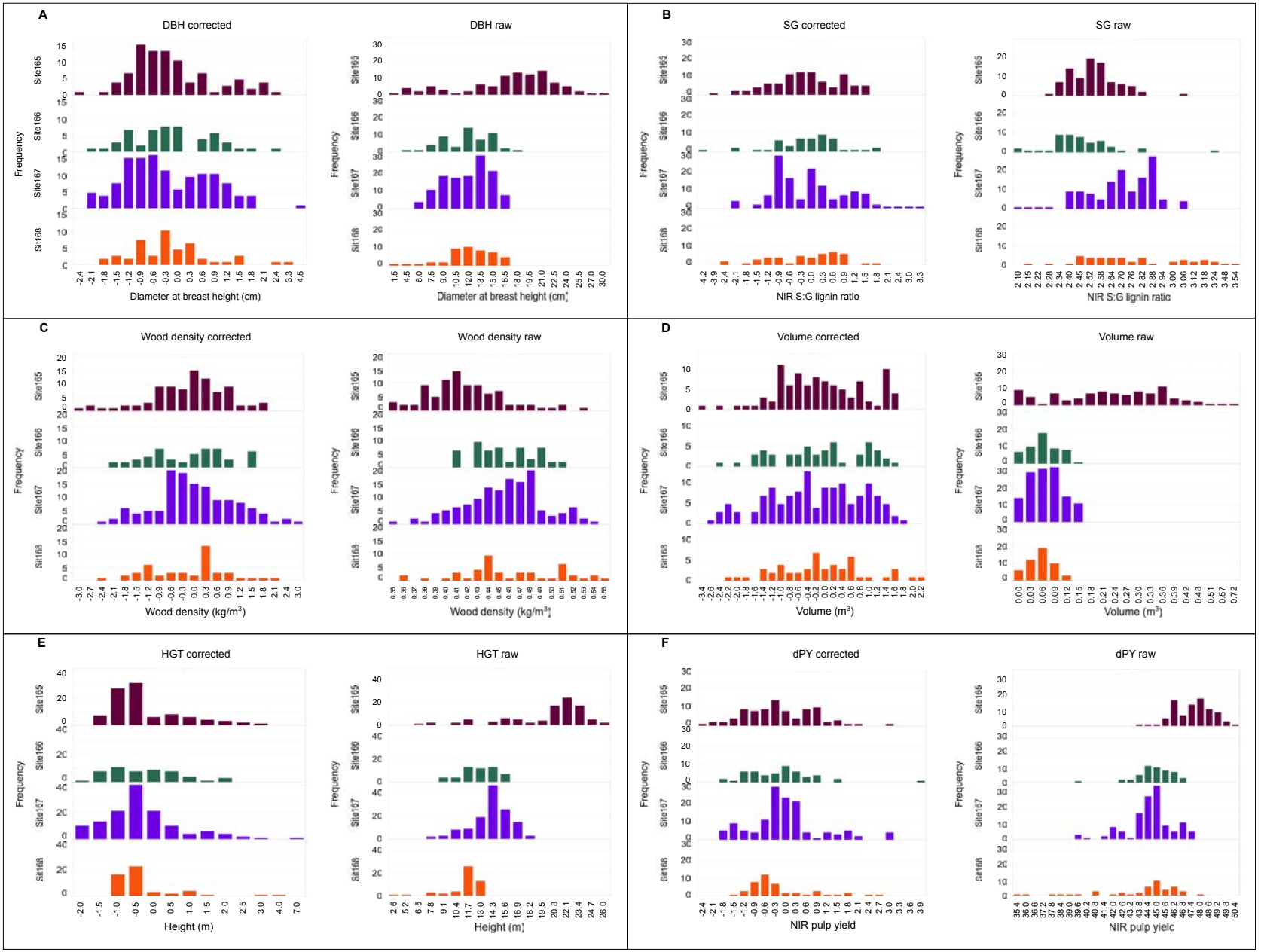


Supplementary Figure 2.5 (*Continued*).



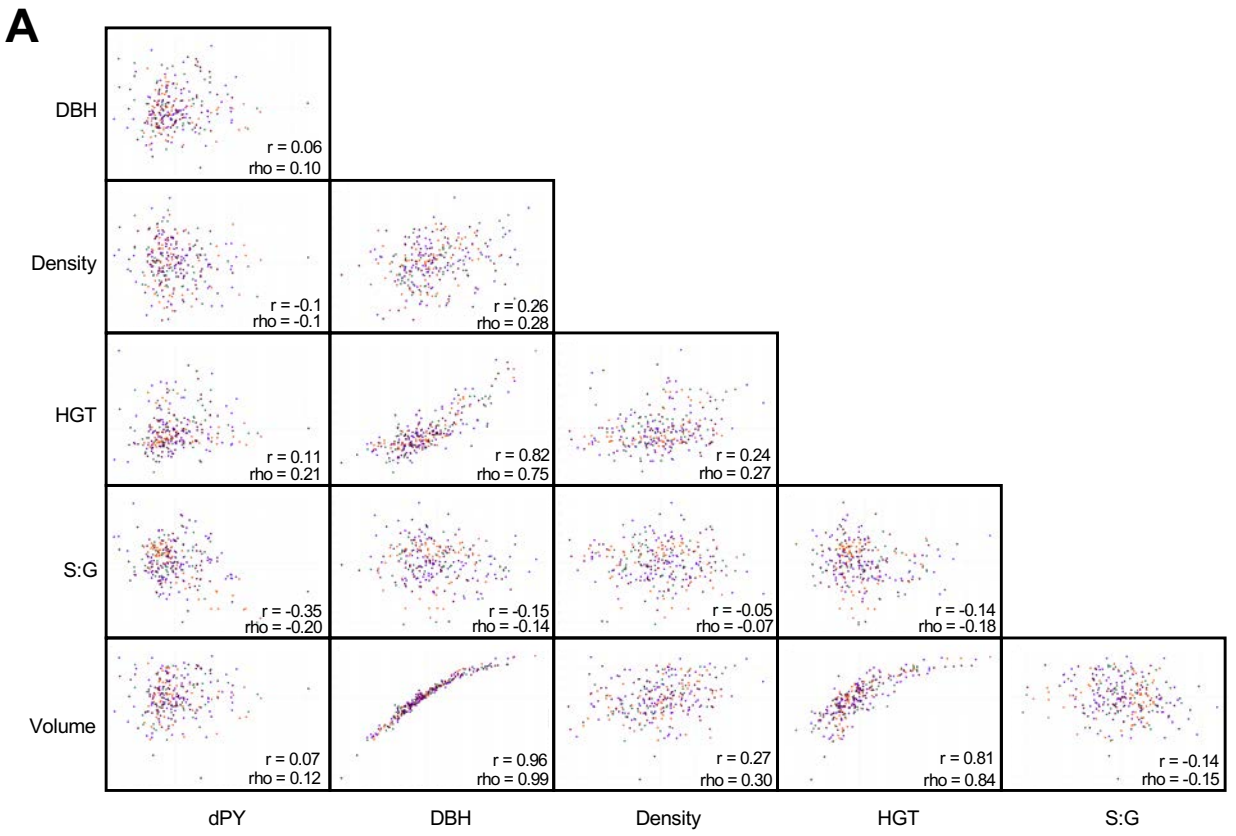
Supplementary Figure 2.6 (Legend on page 84)

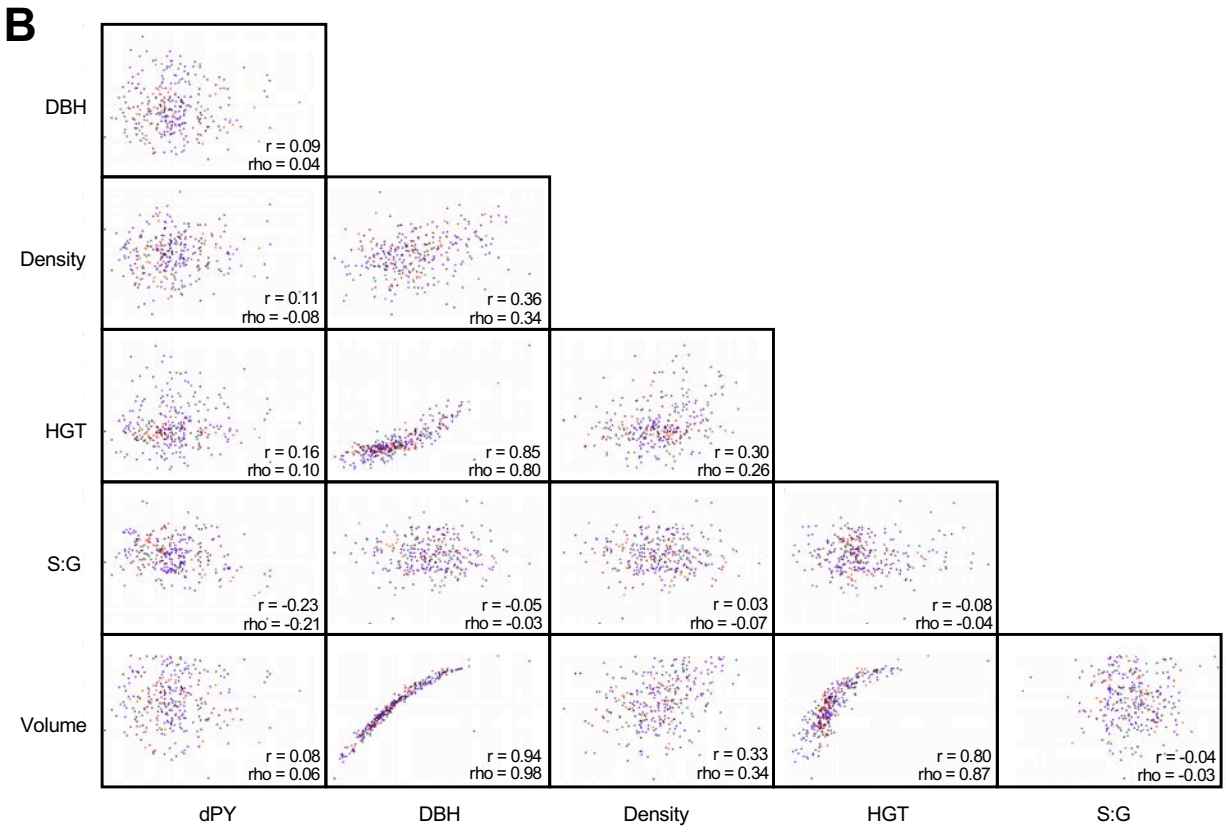
Supplementary Figure 2.6 Trait distribution for the *E. grandis* HS family across the different sites. The x-axis represents the values of each trait while the y-axis represents the frequency (number of individuals). **A.** Diameter at breast height (DBH) **B.** NIR S:G lignin ratio (SG) **C.** Wood density **D.** Volume **E.** Height (HGT) **F.** NIR dissolving pulp yield (dPY). The graphs on the left are for the corrected data ((individual value – mean of site) / SD of site) and the graphs on the right are for the raw data. The mean and SD of the corrected data is 0 and 1 respectively. From the raw data it can be seen that there was a site effect.



Supplementary Figure 2.7 (Legend on page 86)

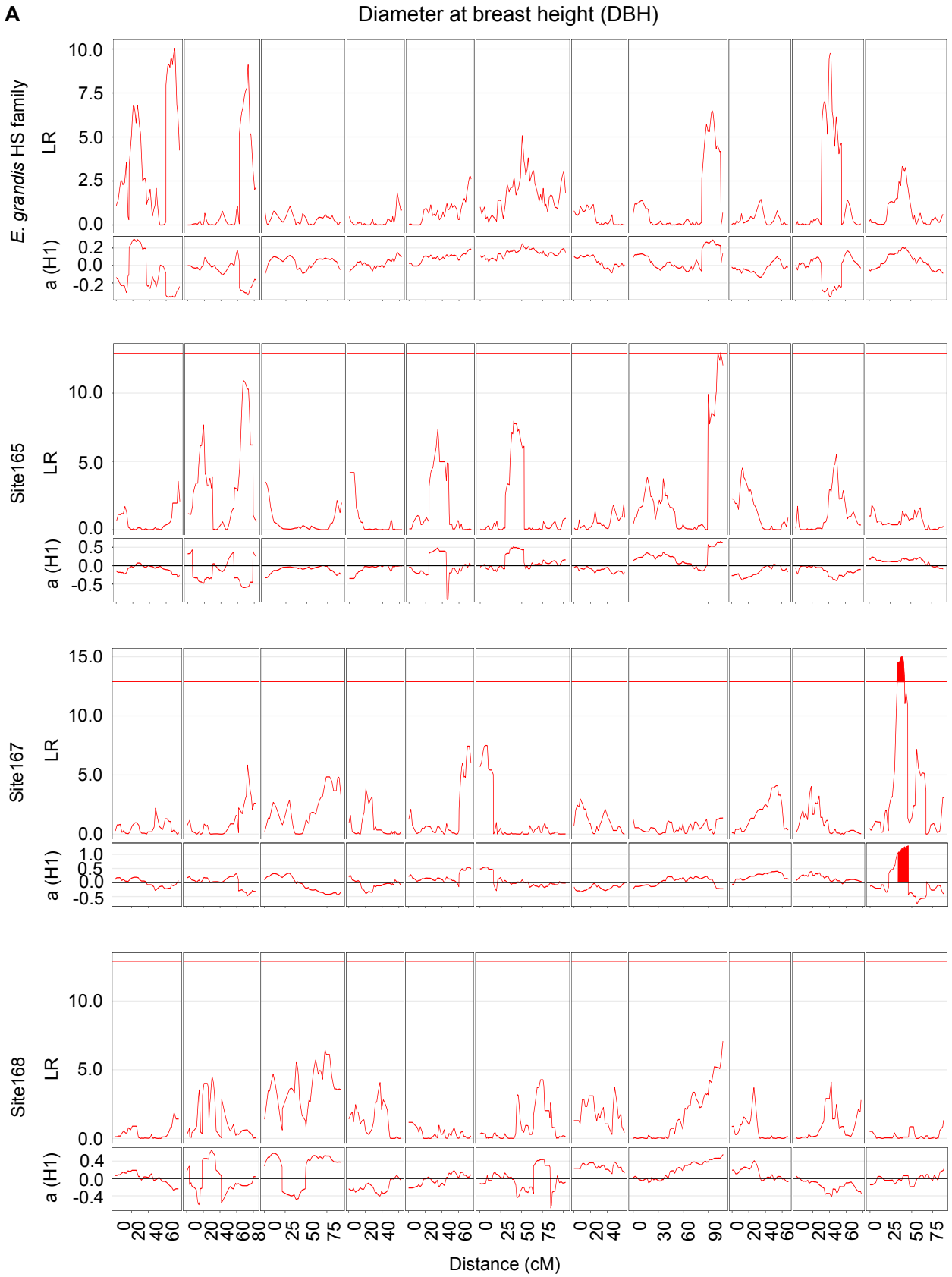
Supplementary Figure 2.7 Trait distribution for the *E. urophylla* HS family across the different sites. The x-axis represents the values of each trait while the y-axis represents the frequency (number of individuals). **A.** Diameter at breast height (DBH) **B.** NIR S:G lignin ratio (SG) **C.** Wood density **D.** Volume **E.** Height (HGT) **F.** NIR dissolving pulp yield (dPY). The graphs on the left are for the corrected data ((individual value – mean of site) / SD of site) and the graphs on the right are for the raw data. The mean and SD of the corrected trait data is 0 and 1 respectively. From the raw data it can be seen that there was a site effect.



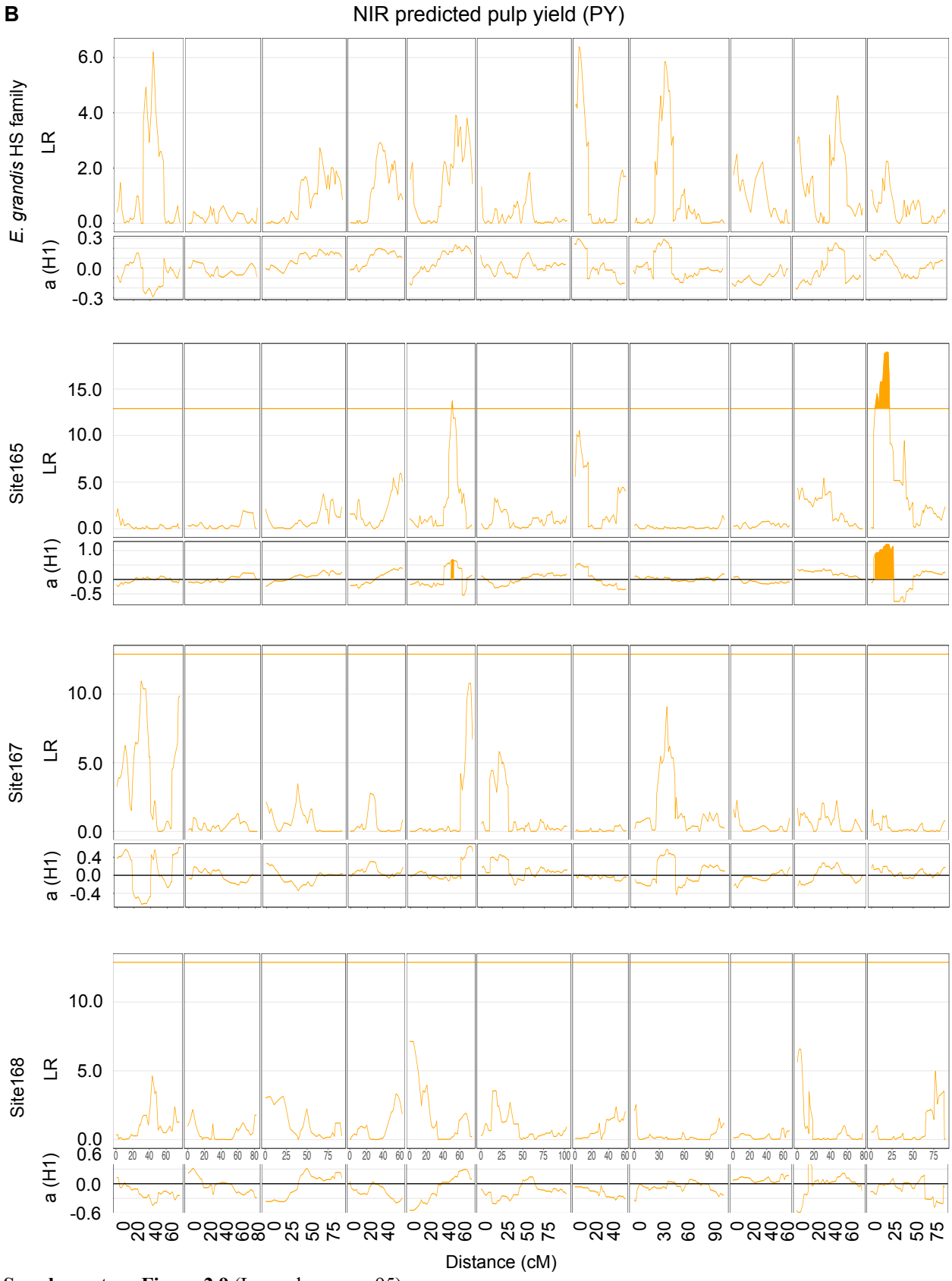


Supplementary Figure 2.8 (Legend on page 88)

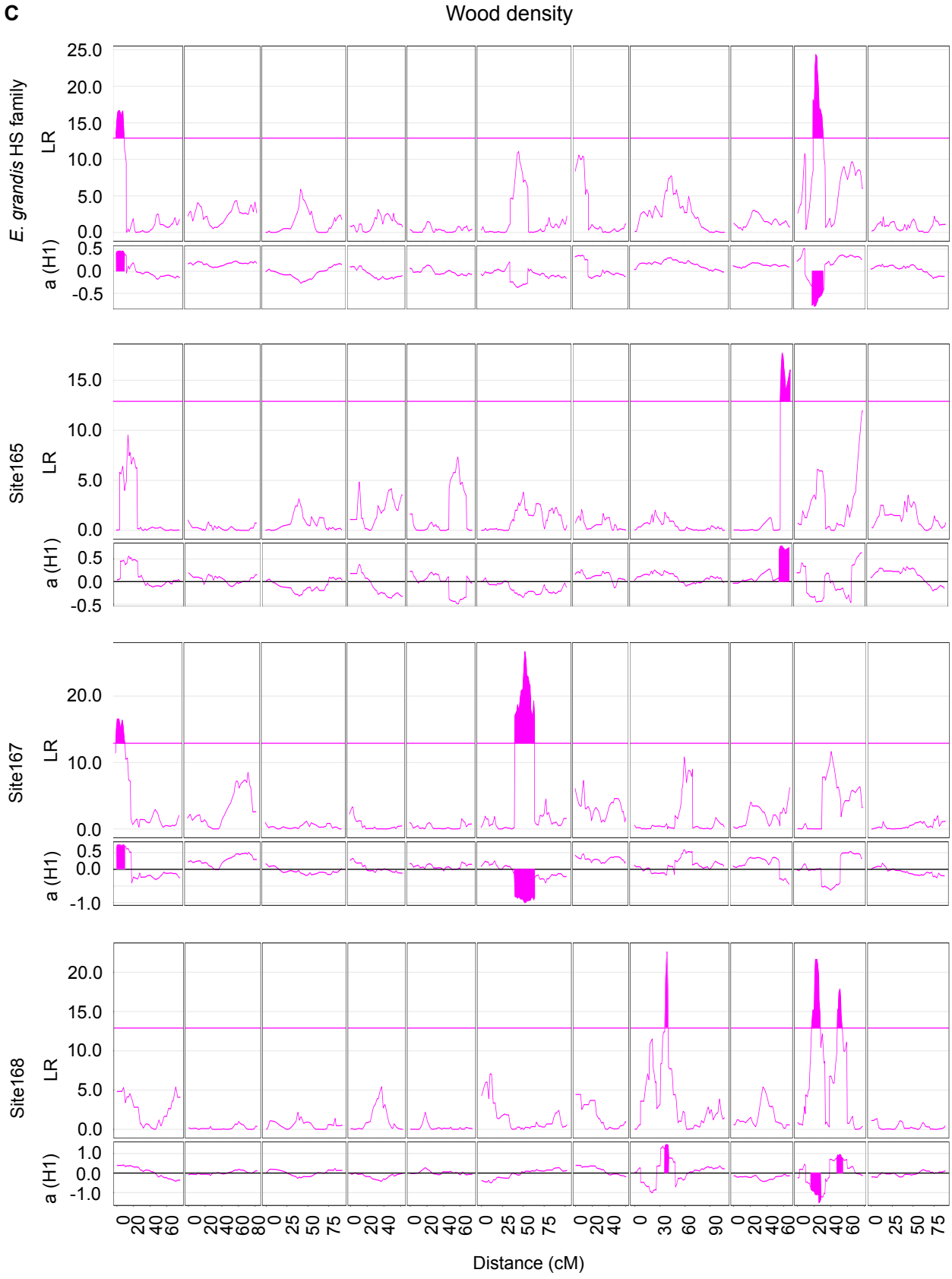
Supplementary Figure 2.8 Phenotypic correlations for all corrected trait data. Correlation performed with Pearson (r) and Spearman (rho) tests for *E. grandis* HS family **(A)** and *E. urophylla* HS family **(B)**. Traits analysed are diameter at breast height (DBH), height (HGT), volume, wood density, NIR dissolving pulp yield (dPY) and NIR S:G lignin ratio (S:G). In both the *E. grandis* and *E. urophylla* HS families, DBH, height and volume are highly correlated.



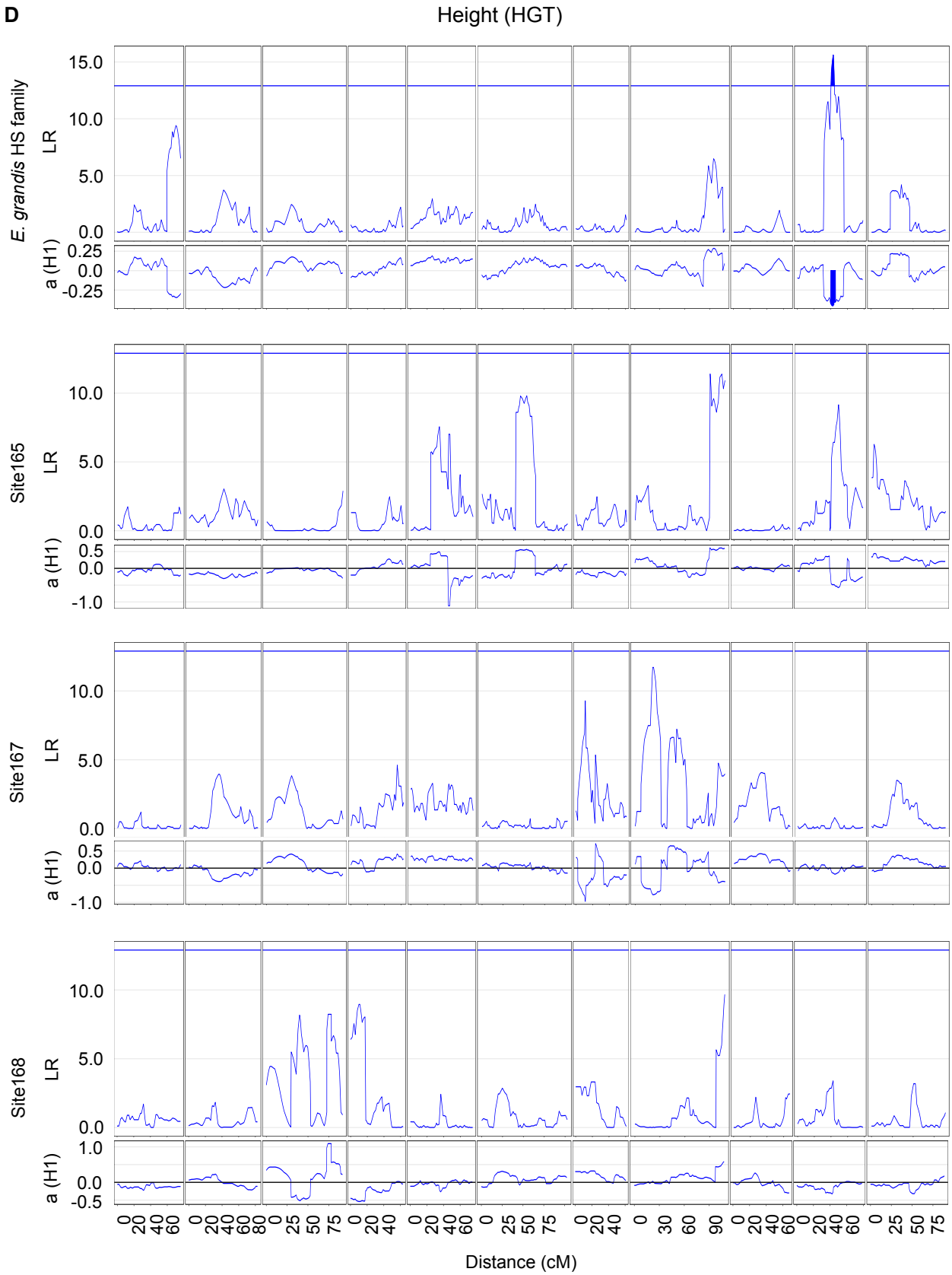
Supplementary Figure 2.9 (Legend on page 95)



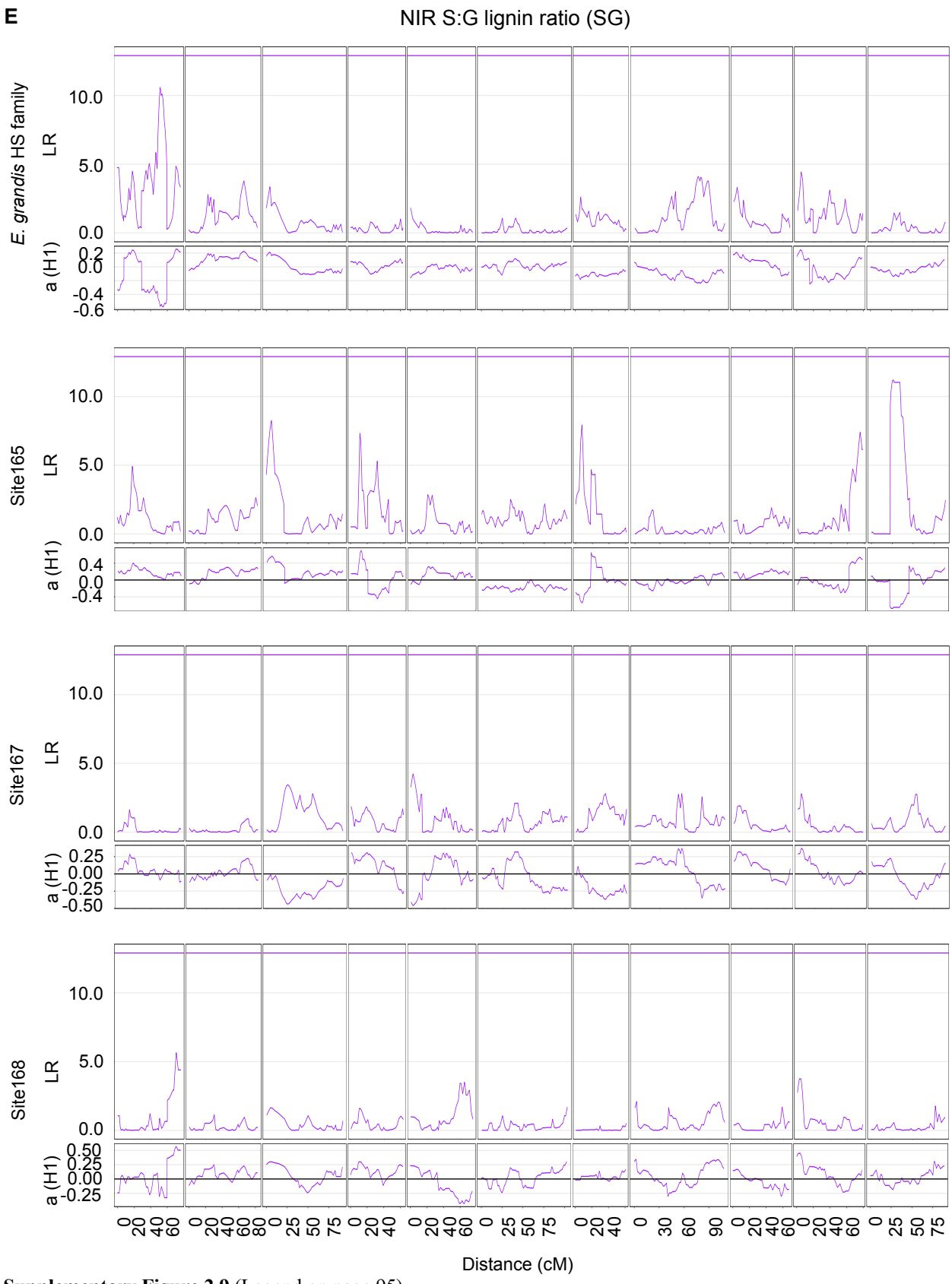
Supplementary Figure 2.9 (Legend on page 95)



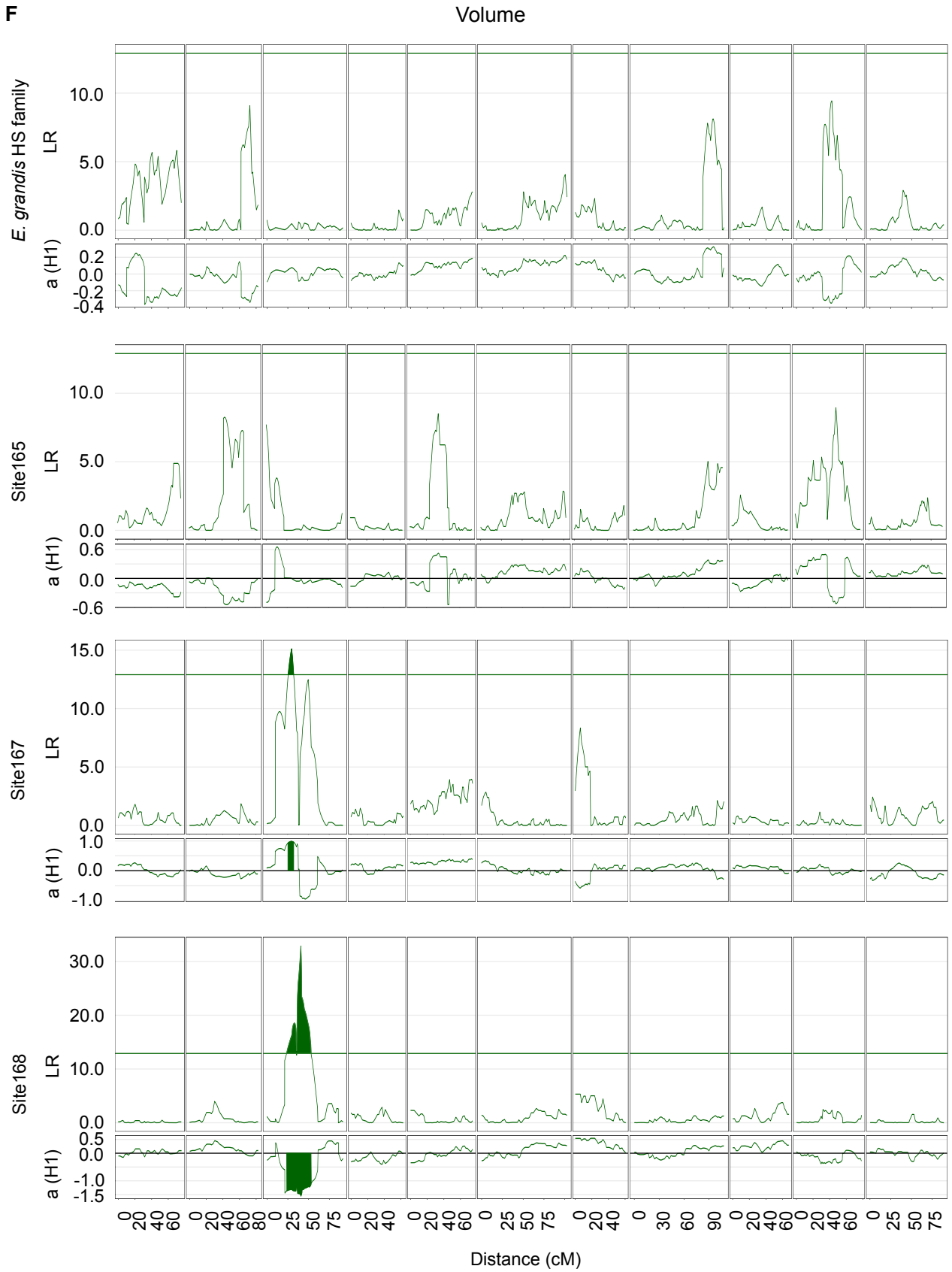
Supplementary Figure 2.9 (Legend on page 95)



Supplementary Figure 2.9 (Legend on page 95)

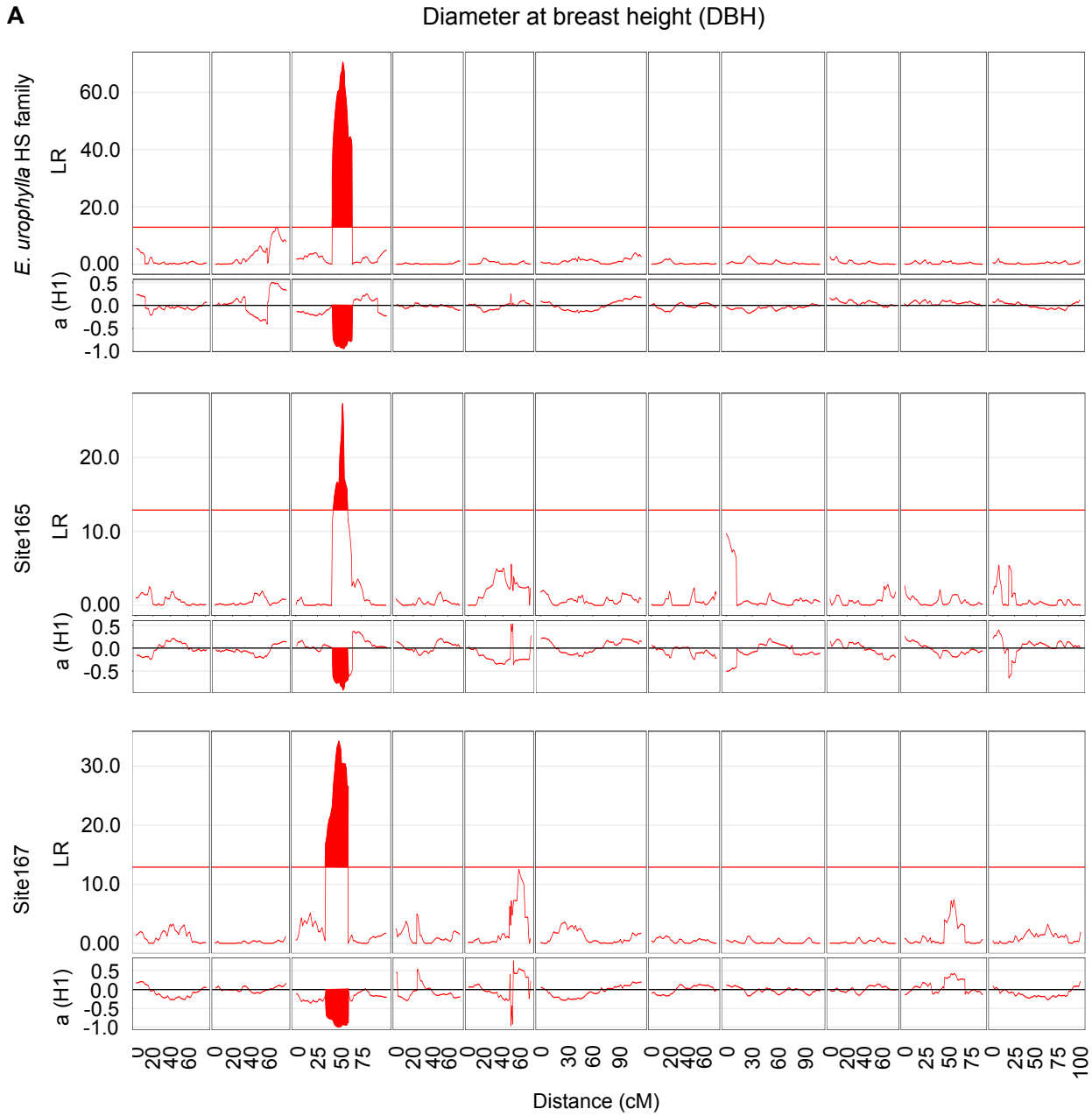


Supplementary Figure 2.9 (Legend on page 95)

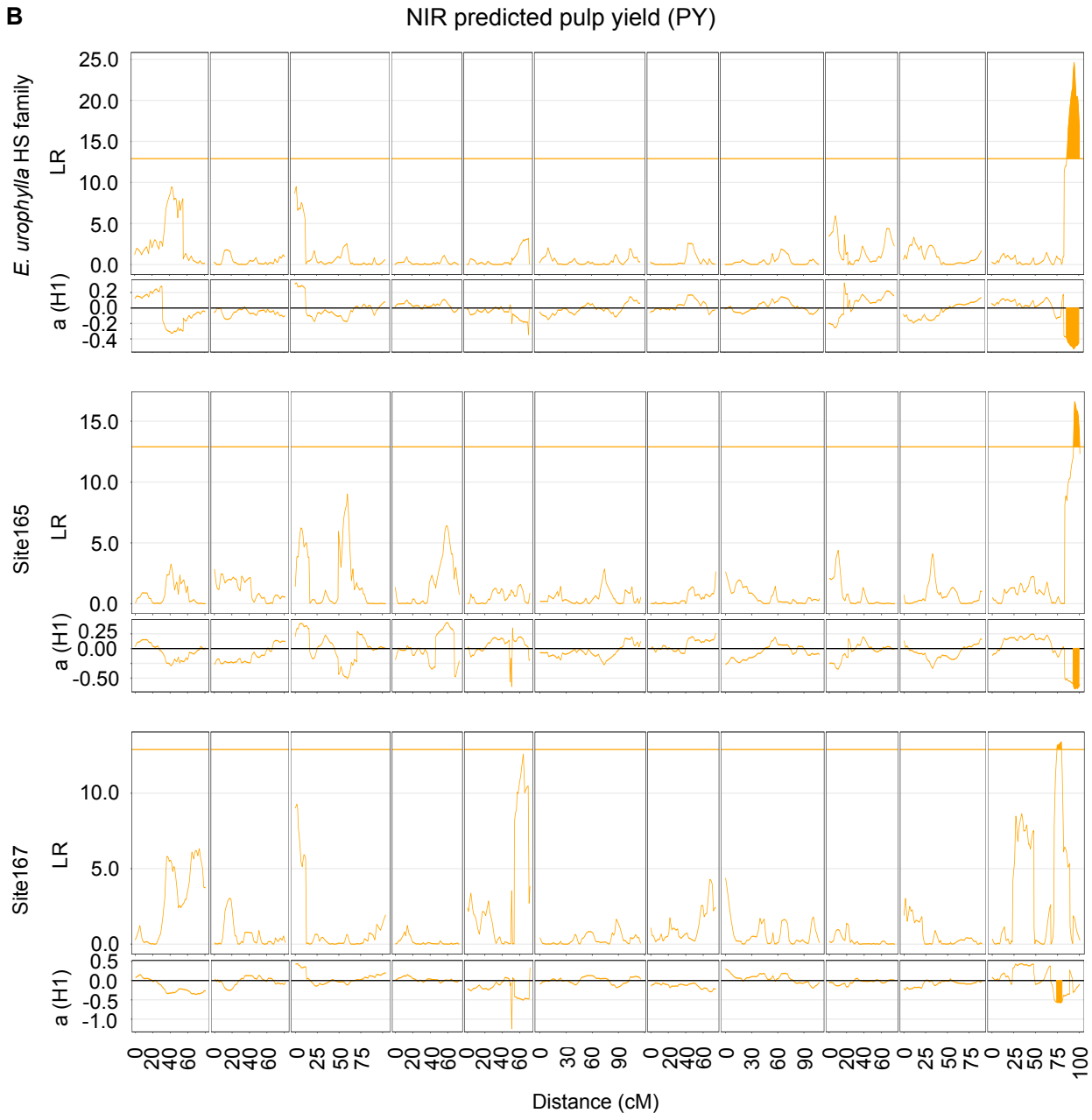


Supplementary Figure 2.9 (Legend on page 95)

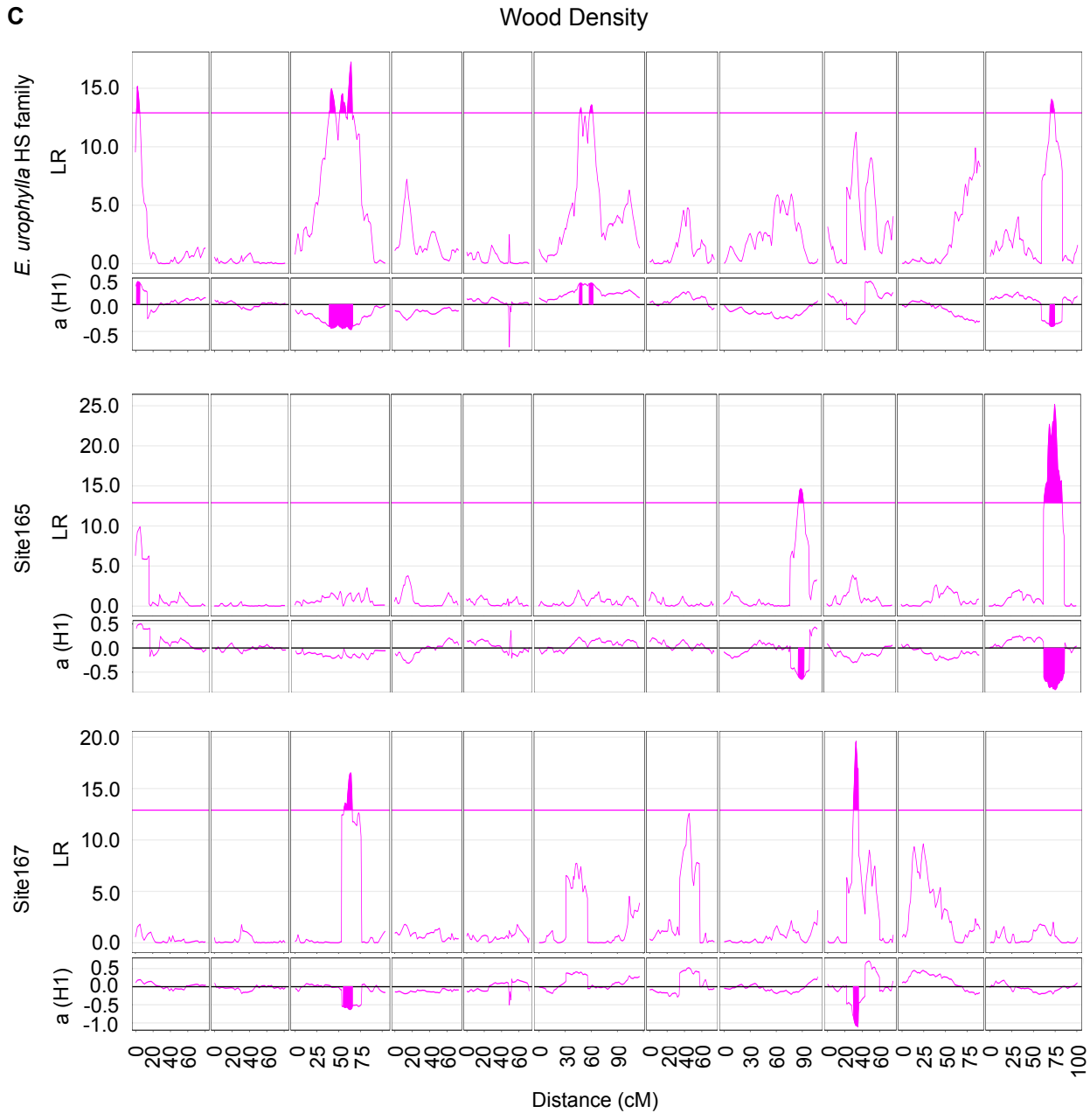
Supplementary Figure 2.9 QTL profiles for *E. grandis* HS family jointly and across three sites. QTLs were identified using QTLCartographer (Basten *et al*. 1994, 2004) and the profiles visualised using R. The experiment-wide LR threshold of 12.9 was determined using a permutation test. The LR values and additive effect values (a (H1)) are shown on the y-axis. The x-axis represents the genome with the positions in centimorgan. Shaded regions indicate QTL peaks that cross the LR threshold. **A.** Diameter at breast height (DBH) **B.** NIR predicted dissolving pulp yield (dPY) **C.** Wood density **D.** Height (HGT) **E.** NIR S:G lignin ratio (SG) **F.** Volume.



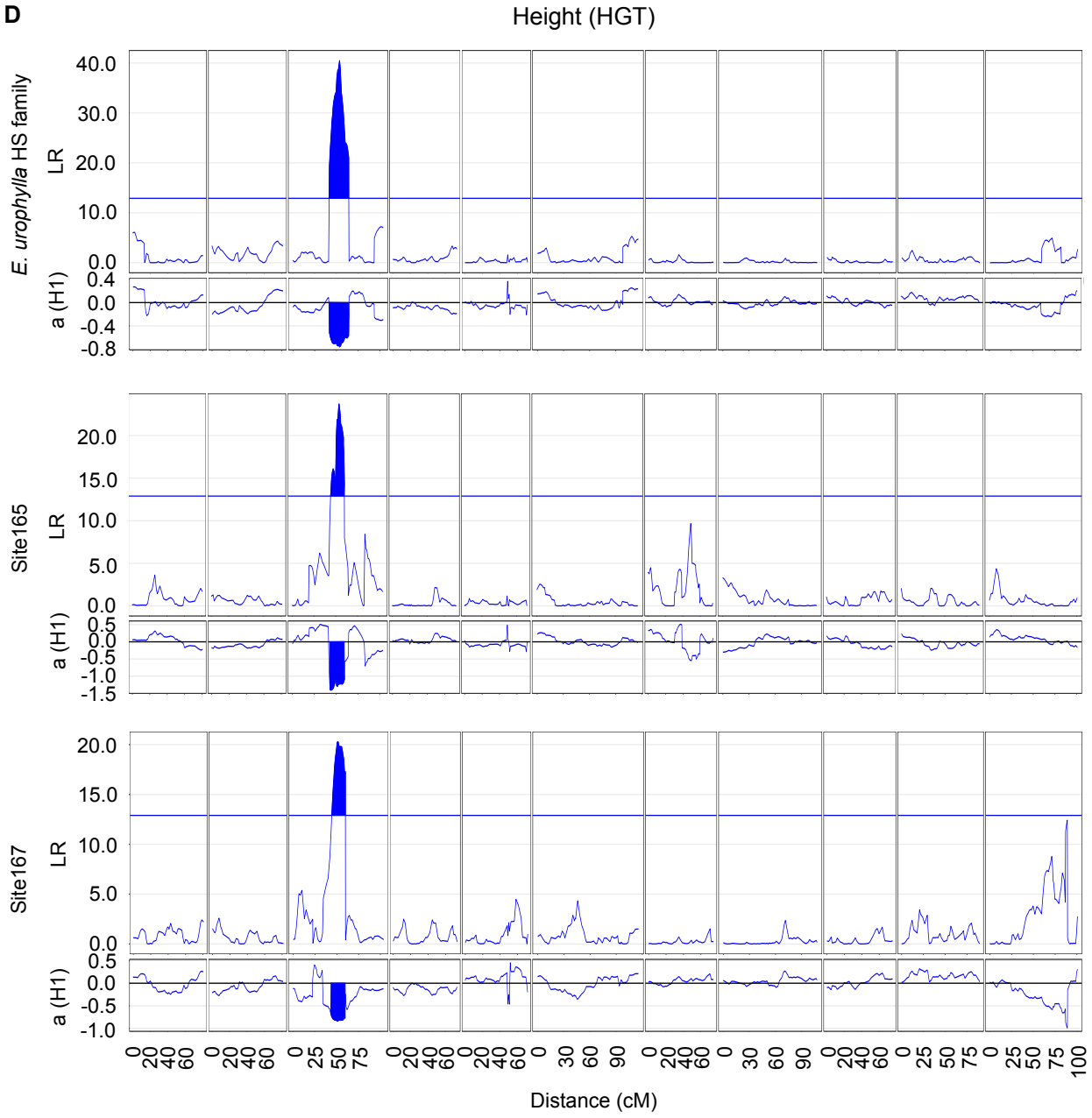
Supplementary Figure 2.10 (Legend on page 102)



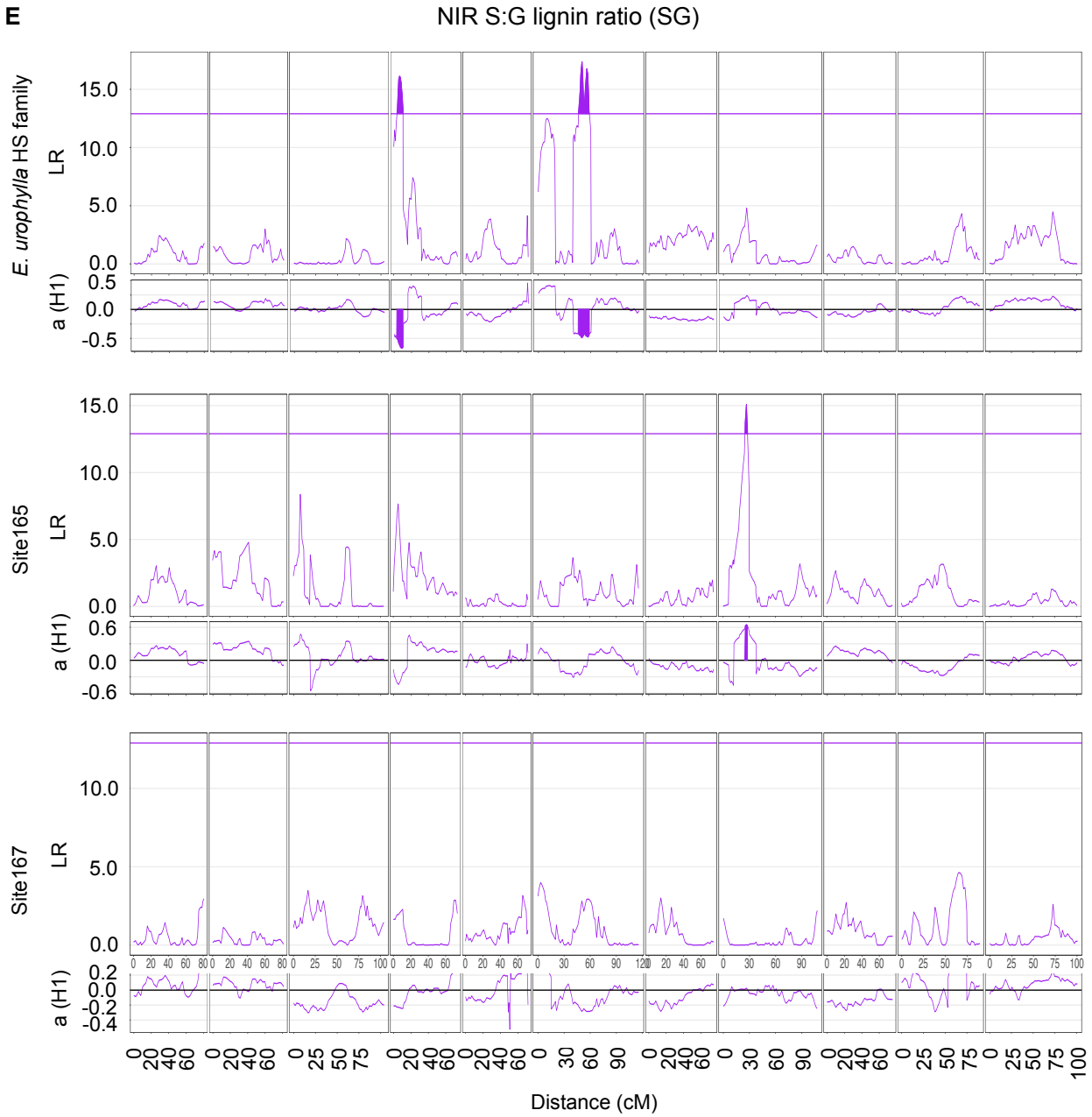
Supplementary Figure 2.10 (Legend on page 102)



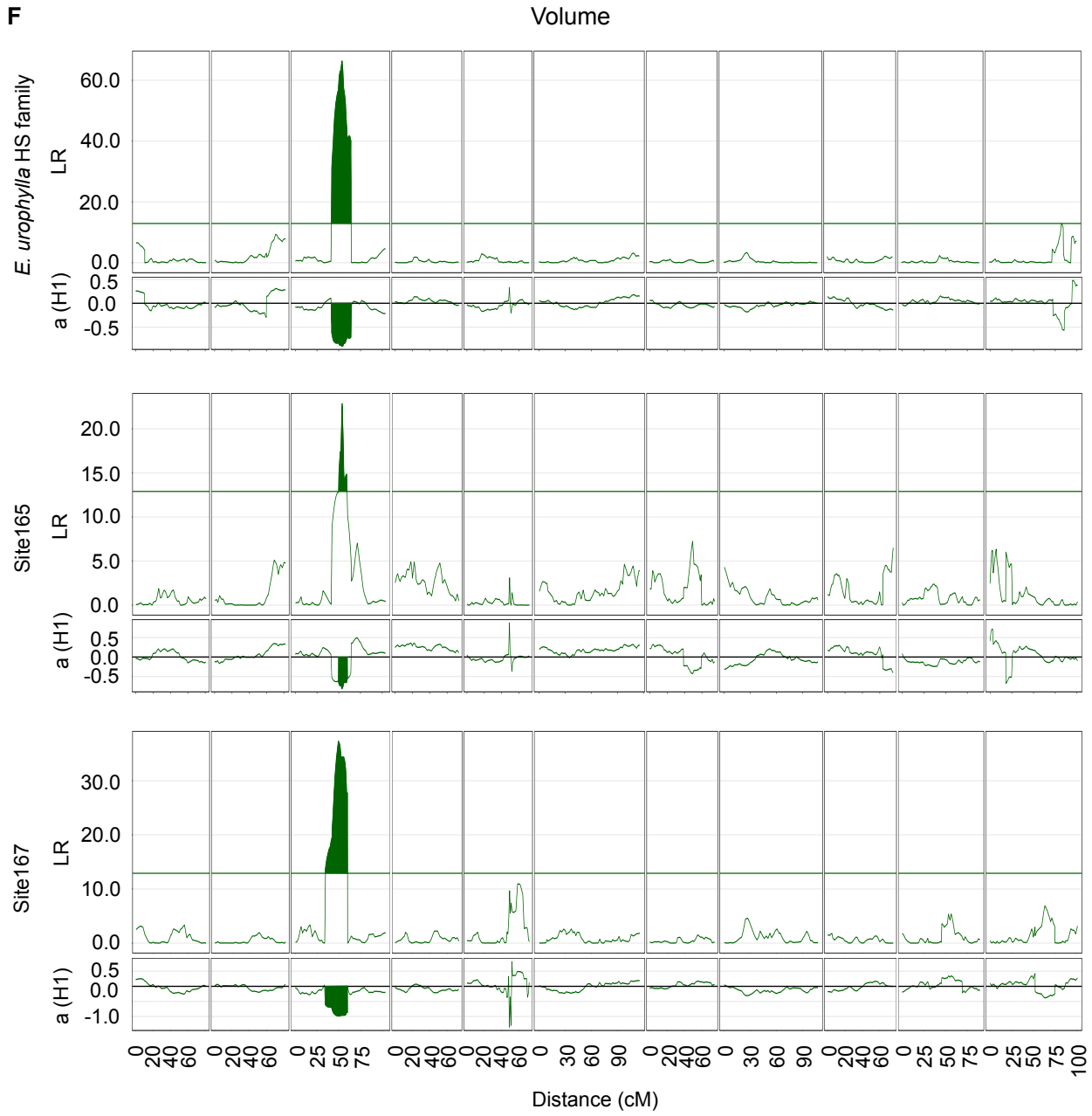
Supplementary Figure 2.10 (Legend on page 102)



Supplementary Figure 2.10 (Legend on page 102)

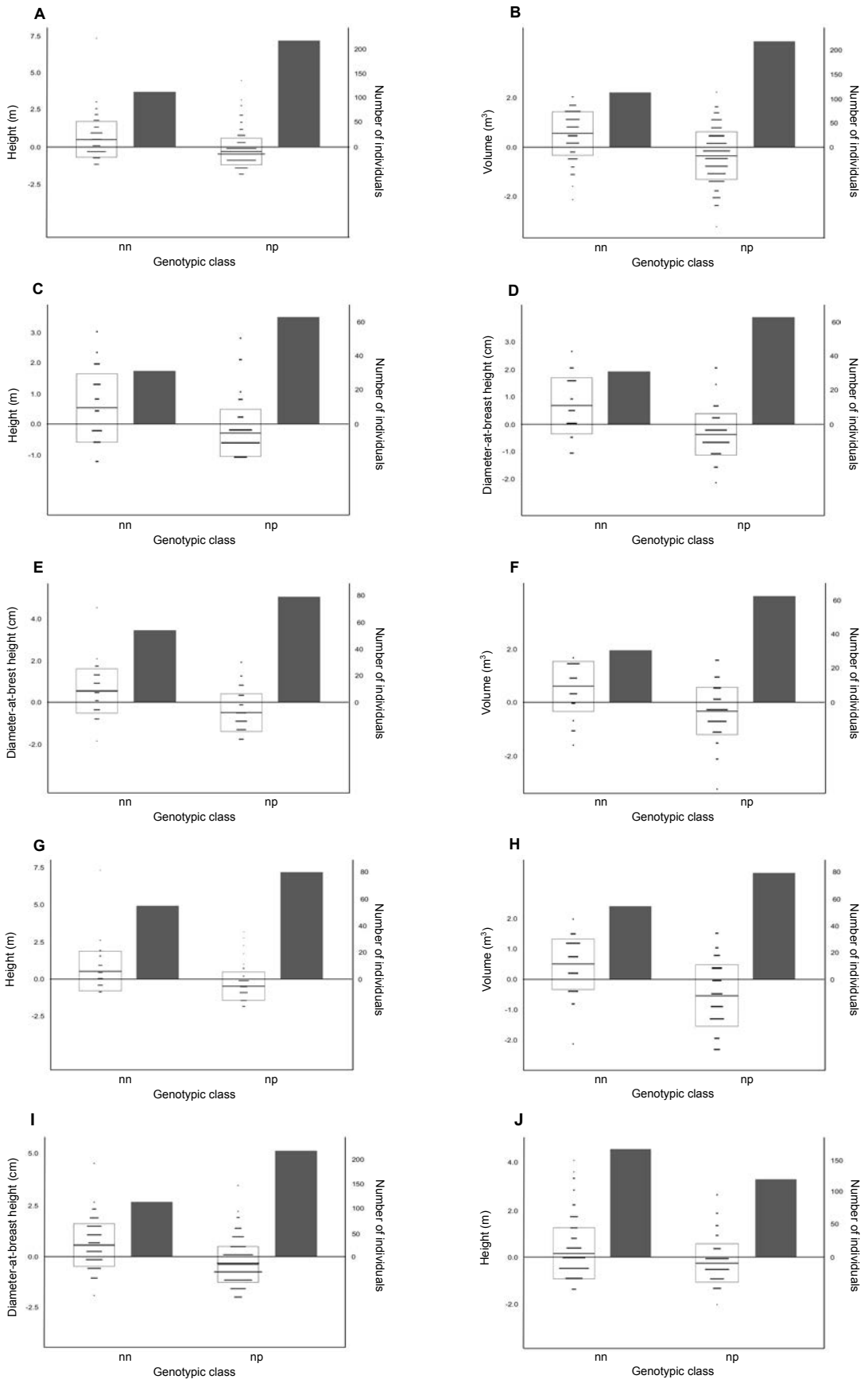


Supplementary Figure 2.10 (Legend on page 102)



Supplementary Figure 2.10 (Legend on page 102)

Supplementary Figure 2.10 QTL profiles for the *E. urophylla* HS family jointly and across two sites. QTLs were identified using QTLCartographer (Basten *et al*. 1994, 2004) and the profiles visualised using R. The genome-wide LR threshold of 12.9 was determined using a permutation test (at $\alpha = 0.05$). The LR values and additive effect values (a (H1)) are shown on the y-axis. The x-axis represents the genome with the positions in centimorgan. The shaded regions indicate QTL peaks which cross the LR threshold. **A.** Diameter at breast height (DBH) **B.** NIR predicted dissolving pulp yield (dPY) **C.** Wood density **D.** Height (HGT) **E.** NIR S:G lignin ratio (SG) **F.** Volume.



Supplementary Figure 2.11 (Legend on page 104)

Supplementary Figure 2.11 Trait data and number of individuals of each genotypic class for growth QTL detected in regions of significant segregation distortion. The dot and box plots represent the standardised trait data for individuals in each genotypic class. Trait data was standardised for site using the following formula; (mean trait value on a site – individual value)/standard deviation of site. Therefore, smaller standardised trait values in the graph correspond to larger actual trait values. The bar plots represent the number of individuals in each genotypic class. **A.** Height QTL detected in the *E. urophylla* HS family on chromosome 3. **B.** Volume QTL detected in the *E. urophylla* HS family on chromosome 3. **C.** Height QTL detected for the *E. urophylla* HS family chromosome 3 on site165. **D.** Diameter at breast height QTL detected for the *E. urophylla* HS family chromosome 3 on site165. **E.** Diameter at breast height QTL detected for the *E. urophylla* HS family chromosome 3 on site167. **F.** Volume QTL detected for the *E. urophylla* HS family chromosome 3 on site165. **G.** Height QTL detected for the *E. urophylla* HS family chromosome 3 on site 167. **H.** Volume QTL detected for the *E. urophylla* HS family chromosome 3 on site167. **I.** Diameter at breast height QTL detected for the *E. urophylla* HS family chromosome 3. **J.** Height QTL detected for the *E. grandis* HS family chromosome 10.

2.12 Supplementary Tables

Supplementary Table 2.1 Site information for the four sites across which the multi-parent population was planted.

Trial number	Location	Land type (SQ)	Altitude (masl)	Latitude	Longitude	Mean Annual Temperature (ºC)	Mean Annual Precipitation (mm)
165	Kwambo Timbers (KT)	003 (I)	60	28º 36' 46.40" S	32º 09' 45.11" E	21.4	1191.4
166	Palm Ridge	002 (III)	39	28º18' 22.79" S	32º 16' 35.10" E	21.8	913.8
167	Clan	113 (II)	822	29º 22' 15.38" S	30º 23' 41.13" E	17.6	1195.3
168	Sabey	123 (III)	978	25º 36' 27.67" S	30º 49' 14.49" E	18.3	951

Supplementary Table 2.2 Colony run settings.

Setting	Criteria
Empirical	Male and Female polygamous, without inbreeding or clones
Species	Monoecious
Length of run	Medium
Analysis	Full likelihood
Likelihood precision	Medium
Update allele frequency	No
Sibship size scaling	Yes
Random number seed and sibship prior	Default
Markers	9
Allele frequency	Unknown
Number of males and females	0
Number of known maternal/paternal sibs	0
Number excluded maternal/paternal sibships	0
Run	Number of threads = 1

Supplementary Table 2.3 Markers which mapped to a different linkage group in the *E. grandis* genetic linkage map compared to *E. grandis* V2 genome assembly. Genetic positions of markers in the full and framework genetic linkage maps were compared with their physical positions in the *E. grandis* v2 reference genome (*https://phytozome.jgi.doe.gov/*).

Locus	Chromosome in *E. grandis* V2 assembly	Linkage group	Present in framework/full/both genetic linkage maps
EuBR03s16773516	3	1	Full
EuBR01s1570286	7	1	Full
EuBR08s707846	8	2	Full
EuBR06s23537926	6	2	Full
EuBR02s12531252	4	2	Full
EuBR07s19312362	7	2	Full
EuBR02s32599377	6	2	Full
EuBR03s53640626	5	3	Full
EuBR06s991307	6	3	Full
EuBR06s1421821	6	3	Full
EuBR06s1152592	6	3	Full
EuBR02s41739709	2	5	Full
EuBR03s61635633	3	5	Both
EuBR04s3782009	4	5	Full
EuBR04s11810383	4	5	Full
EuBR03s1472848	3	5	Full
EuBR04s41704829	4	5	Both
EuBR03s79886124	3	5	Full
EuBR05s74058904	5	6	Full
EuBR07s52199270	3	7	Full
EuBR07s52232983	3	7	Both
EuBR07s46525870	11	7	Full
EuBR07s3765369	8	7	Full
EuBR03s5400509	3	7	Full
EuBR08s55309610	8	7	Full
EuBR07s49302025	7	8	Both
EuBR05s37951602	5	8	Full
EuBR04s24468179	8	9	Full
EuBR08s5215966	8	10	Full
EuBR08s5217929	8	10	Full
EuBR06s47505963	6	10	Full
EuBR03s37188589	3	11	Full

Supplementary Table 2.4 *E. urophylla* markers which mapped to a different linkage group compared to the chromosome position in the *E. grandis* V2 reference genome assembly. Genetic positions of markers in the full and framework genetic linkage maps were compared with their physical positions in the *E. grandis* v2 reference genome (*https://phytozome.jgi.doe.gov/*).

Locus	Chromosome in *E. grandis* V2 assembly	Linkage group	Present in framework/full/both genetic linkage maps
EuBR08s39496756	8	2	Full
EuBR06s1297770	6	3	Full
EuBR06s1480294	6	3	Full
EuBR04s34499660	2	4	Full
EuBR11s10662569	11	5	Both
EuBR07s12765217	11	5	Full
EuBR05s23227843	5	6	Full
EuBR03s13990906	5	6	Full
EuBR07s15234370	7	6	Both
EuBR10s14434683	10	6	Both
EuBR02s52970568	2	7	Full
EuBR03s17960592	3	7	Full
EuBR04s28655644	4	7	Full
EuBR08s41408832	8	7	Full
EuBR08s68713843	8	7	Full
EuBR08s57907223	11	7	Full
EuBR01s29775581	1	8	Full
EuBR06s19317853	6	8	Full
EuBR07s171841	7	9	Full
EuBR05s3053916	5	11	Full

Supplementary Table 2.5 Summary statistics of the raw trait data of the *E. grandis* HS family. Traits analysed were diameter at breast height (DBH), height, volume, wood density, NIR dissolving pulp yield (dPY) and NIR S:G lignin ratio (S:G). Mean, standard deviation (SD) minimum (min) and maximum values were calculated for each trait in the *E. grandis* HS family and across the four sites. Shapiro-Wilk test for normality was performed for each trait.

E. grandis HS family raw data (n = 349)						
					Shapiro-Wilk W Test	
Trait	Mean	SD	Min	Max	W	p-value
DBH	14.01	4.07	2.10	30.00	0.99	0.02*
Height	15.42	4.11	3.20	25.20	0.97	3.55E-06*
Volume	0.12	0.10	0.00	0.72	0.81	<2.00e-16*
Wood density	0.44	0.04	0.34	0.55	0.99	0.12
dPY	45.50	2.04	37.34	50.16	0.96	1.53E-06*
SG	2.67	0.26	2.07	3.73	0.95	3.13E-08*
E. grandis Site165 raw data (n = 105)						
DBH	16.99	4.85	4.50	30.00	0.96	4.00E-03*
Height	20.06	3.36	8.20	25.20	0.87	4.57E-07*
Volume	0.21	0.12	0.01	0.72	0.94	9.41E-04*
Wood density	0.41	0.04	0.35	0.50	0.97	0.08
dPY	47.50	1.26	41.08	50.16	0.90	6.43E-06*
SG	2.58	0.21	2.07	3.19	1.00	0.86
E. grandis Site166 raw data (n = 62)						
DBH	12.20	2.85	6.50	17.50	0.04	0.02*
Height	213.20	1.81	9.90	16.80	0.96	0.12
Volume	0.07	0.03	0.01	0.14	0.97	0.22
Wood density	0.46	0.03	0.38	0.52	0.96	0.09
dPY	44.57	1.47	39.94	46.81	0.92	3.00E-03*
SG	2.58	0.22	2.16	3.27	0.93	0.01*
E. grandis Site167 raw data (n = 99)						
DBH	13.20	2.59	6.00	18.00	0.96	0.01*
Height	14.99	1.79	8.50	17.70	0.90	9.85E-06*
Volume	0.09	0.04	0.01	0.17	0.98	0.19
Wood density	0.46	0.04	0.34	0.55	0.98	0.17
dPY	44.68	1.22	41.84	47.49	0.95	6.00E-03*
SG	2.74	0.20	2.15	3.11	0.96	0.03*
E. grandis Site168 raw data						
DBH	12.48	3.01	2.10	17.50	0.94	2.00E-03*
Height	11.67	2.06	3.20	13.40	0.73	5.67E-10*
Volume	0.06	0.03	0.00	0.13	0.99	0.78
Wood density	1.45	0.05	0.35	0.55	0.97	0.16
dPY	44.57	2.19	37.34	47.12	0.79	2.79E-08*
SG	2.77	0.34	2.35	3.73	0.84	5.47E-07*

* Statistically significant at $P < 0.05$

Supplementary Table 2.6 Summary statistics of the raw trait data of the *E. urophylla* HS family. Traits analysed were diameter at breast height (DBH), height, volume, wood density, NIR dissolving pulp yield (dPY) and NIR S:G lignin ratio (S:G). Mean, standard deviation (SD) minimum (min) and maximum values were calculated for each trait in the *E. urophylla* HS family and across the four sites. Shapiro-Wilk test for normality was performed for each trait.

E. urophylla HS family raw data (n = 367)						
						Shapiro-Wilk W test
Trait	Mean	SD	Min	Max	W	p-value
DBH	14.08	4.53	2.10	30.00	0.98	1.00E-03*
Height	15.78	4.24	3.20	26.10	0.94	1.36E-10*
Volume	0.13	0.11	0.00	0.72	0.80	<2.20E-16*
Wood density	0.45	0.04	0.35	0.57	1.00	0.68
dPY	45.60	2.21	35.83	50.59	0.95	3.94E-09*
SG	2.66	0.23	2.14	3.57	0.97	1.50E-05*
E. urophylla Site165 raw data (n = 100)						
DBH	17.66	5.77	2.30	30.00	0.94	4.58E-04*
Height	20.51	4.27	7.70	26.10	0.85	1.83E-08*
Volume	0.25	0.15	0.01	0.72	0.97	0.02*
Wood density	0.42	0.04	0.35	0.53	0.97	0.04*
dPY	47.78	1.28	43.81	50.59	0.99	0.43
SG	2.57	0.13	2.34	3.07	0.97	0.02*
E. urophylla Site166 raw data (n = 67)						
DBH	12.55	2.95	5.20	18.00	0.97	0.23
Height	13.42	1.74	9.20	16.80	0.98	0.37
Volume	0.07	0.04	0.01	0.16	0.97	0.15
Wood density	0.46	0.03	0.41	0.52	0.96	0.06
dPY	45.15	1.25	39.94	47.24	0.91	9.13E-04*
SG	2.49	0.18	2.15	3.25	0.90	5.58E-04*
E. urophylla Site167 raw data (n = 141)						
DBH	12.63	2.76	6.70	17.60	0.97	0.01*
Height	14.79	2.06	8.30	18.50	0.95	1.98E-04*
Volume	0.08	0.04	0.01	0.17	0.97	0.01*
Wood density	0.46	0.04	0.35	0.55	0.99	0.47
dPY	44.83	1.58	39.95	47.77	0.92	2.08E-06*
SG	2.73	0.19	2.14	3.11	0.95	1.86E-04*
E. urophylla Site168 raw data(n = 59)						
DBH	12.77	3.07	2.10	17.50	0.93	0.01*
Height	11.92	1.95	3.20	13.40	0.65	1.03E-09*
Volume	0.07	0.03	0.00	0.13	0.98	0.75
Wood density	0.47	0.05	0.36	0.57	0.98	0.44
dPY	44.08	2.91	35.83	48.04	0.85	4.16E-05*
SG	2.81	0.32	2.17	3.57	0.96	0.14

* Statistically significant at $P < 0.05$

Supplementary Table 2.7 Two-way ANOVA of phenotypic data for the *E. grandis* HS family. The following traits were analysed; diameter at breast height (DBH), height, volume, wood density, NIR dissolving pulp yield (dPY) and S:G lignin ratio (S:G).

E. grandis HS (n = 349)					
Variable	df	SumSq	MeanSq	F-value	Pr(>F)
Diameter at breast height (DBH)					
Family	1	84.00	84.40	2.03	0.16
Site	1	612.00	612.00	14.69	1.50E-04*
Family and site	1	6.00	6.00	0.14	0.70
Height					
Family	1	112.00	112.10	2.53	0.11
Site	1	1961.00	1961.40	44.28	1.12E-10
Family and site	1	1.00	1.40	0.03	0.86
Volume					
Family	1	0.02	0.02	2.66	0.10
Site	1	0.68	0.68	85.65	<2e-16*
Family and site	1	0.00	0.00	0.04	0.85
Wood density					
Family	1	0.11	0.11	3.20	0.07
Site	1	0.02	0.02	0.70	0.40
Family and site	1	0.01	0.01	0.40	0.53
Dissolving pulp yield (dPY)					
Family	1	159.00	158.70	0.47	0.49
Site	1	509.00	509.00	1.51	0.22
Family and site	1	40.00	40.10	0.12	0.73
S:G lignin ratio					
Family	1	1.70	1.71	1.42	0.24
Site	1	0.80	0.83	0.69	0.41
Family and site	1	0.00	0.00	0.00	0.99

* Statistically significant at $P < 0.05$

Supplementary Table 2.8 One-way ANOVA of the phenotypic data, with FS family as the condition tested for the *E. grandis* HS family across sites. The following traits were analysed; diameter at breast height (DBH), height, volume, wood density, NIR dissolving pulp yield (dPY) and NIR S:G ratio (S:G). Significance was tested at a 0.05 threshold. Family did not affect the means across the sites.

Trait	Variable	df	Sum Sq	Mean Sq	F-value	Pr(>F)
E. grandis HS site165 (n = 105)						
DBH	Family	1	9.00	8.89	0.15	0.70
Height	Family	1	20.00	19.68	0.29	0.59
Volume	Family	1	0.01	0.01	0.61	0.44
Density	Family	1	0.02	0.02	0.57	0.45
dPY	Family	1	2.00	1.80	0.01	0.94
S:G	Family	1	0.76	0.76	0.71	0.40
E. grandis HS site166 (n = 62)						
DBH	Family	1	37.90	37.92	1.21	0.28
Height	Family	1	29.00	28.96	0.91	0.34
Volume	Family	1	6.2E-03	6.2E-03	3.76	0.06
Density	Family	1	3.0E-04	3.2E-04	0.01	0.93
dPY	Family	1	0.00	0.30	1.0E-03	0.98
S:G	Family	1	0.00	4.6E-03	4.0E-03	0.95
E. grandis HS site167 (n = 99)						
DBH	Family	1	28.00	27.85	0.83	0.37
Height	Family	1	103.00	103.24	2.68	0.11
Volume	Family	1	3.1E-03	3.1E-03	1.30	0.26
Density	Family	1	0.07	0.07	1.91	0.17
dPY	Family	1	394.00	393.70	1.13	0.29
S:G	Family	1	1.62	1.62	1.21	0.27
E. grandis HS site168 (n = 83)						
DBH	Family	1	27.00	27.05	0.91	0.34
Height	Family	1	10.00	10.01	0.44	0.51
Volume	Family	1	3.4E-03	3.4E-03	2.55	0.12
Density	Family	1	0.05	0.05	1.32	0.25
dPY	Family	1	16.00	15.60	0.05	0.83
S:G	Family	1	0.33	0.33	0.24	0.63

Supplementary Table 2.9 Two-way ANOVA of raw data for the *E. urophylla* HS family using family, site and family within site as the conditions. The following traits were tested; diameter at breast height (DBH), height, volume, wood density, NIR dissolving pulp yield (dPY) and S:G lignin ratio (S:G).

E. urophylla HS (n = 367)					
Variable	df	SumSq	MeanSq	Fvalue	Pr(>F)
Diameter at breast height (DBH)					
Family	1	540.00	89.90	3.05	6.47E-03*
Site	1	2222.00	740.70	25.08	1.06E-14*
Family and site	1	428.00	32.90	1.12	0.34
Height					
Family	1	1178.00	196.30	7.74	7.99E-08*
Site	1	3915.00	1304.90	51.47	2.00E-16*
Family and site	1	388.00	29.80	1.18	0.30
Volume					
Family	1	0.14	0.02	2.90	9.13E-03*
Site	1	1.98	0.66	83.81	2.00E-16*
Family and site	1	0.07	0.01	0.69	0.78
Wood density					
Family	1	0.70	0.12	5.04	5.74E-05*
Site	1	0.25	0.08	3.64	1.31E-02*
Family and site	1	0.68	0.05	2.25	7.83E-03*
Dissolving pulp yeild (dPY)					
Family	1	8188.00	1364.60	5.78	9.55E-06*
Site	1	3436.00	1145.40	4.85	2.56E-03*
Family and site	1	5876.00	452.00	1.91	2.77E-02*
S:G lignin ratio					
Family	1	31.74	5.29	6.66	1.11E-06*
Site	1	14.75	4.92	6.19	4.18E-04*
Family and site	1	27.03	2.08	2.62	1.73E-03*

* Statistically significant at $P < 0.05$

Supplementary Table 2.10 One-way ANOVA for phenotypic data of the *E. urophylla* HS family across site. Family was the condition used and the significance threshold was 0.05. The following traits were analysed; diameter at breast height (DBH), height, volume, wood density, NIR dissolving pulp yield (dPY) and NIR S:G lignin ratio (S:G).

Trait	Variable	df	SumSq	MeanSq	F-value	Pr(>F)
E. urophylla site165 (n = 100)						
DBH	Family	1	131	26.19	0.55	0.74
Height	Family	1	172	34.42	0.91	0.48
Volume	Family	1	0.08	1.59E-02	0.67	0.65
Density	Family	1	0.10	2.07E-02	1.02	0.41
dPY	Family	1	1457	291.40	1.20	0.31
S:G	Family	1	4.54	0.91	1.28	0.28
E. urophylla site166 (n = 67)						
DBH	Family	1	64.50	21.51	0.64	0.59
Height	Family	1	84.70	28.22	0.86	0.47
Volume	Family	1	0.00	0.00	0.20	0.90
Density	Family	1	0.16	0.05	1.35	0.27
dPY	Family	1	1016	338.6	0.82	0.49
S:G	Family	1	2.39	0.80	0.62	0.60
E. urophylla site167 (n = 141)						
DBH	Family	1	282	47.00	3.05	7.84E-03*
Height	Family	1	389.9	64.99	4.33	4.99E-04*
Volume	Family	1	0.01	0.00	1.09	0.37
Density	Family	1	0.53	0.09	7.38	7.80E-07*
dPY	Family	1	4476	745.90	6.16	1.00E-05*
S:G	Family	1	16.57	2.76	5.80	2.14E-05*
E. urophylla site168 (n = 59)						
DBH	Family	1	220.1	44.01	1.57	0.19
Height	Family	1	166.8	33.35	1.60	0.18
Volume	Family	1	0.01	0.00	1.25	0.30
Density	Family	1	0.43	0.09	2.30	0.06
dPY	Family	1	4774	954.80	3.10	1.59E-02*
S:G	Family	1	26.85	5.37	4.60	1.47E-03*

* Statistically significant at $P < 0.05$

Supplementary Table 2.11 Two-way ANOVA of corrected phenotypic data for the *E. grandis* HS family. The following traits were analysed; diameter at breast height (DBH), height, volume, wood density, NIR dissolving pulp yield (dPY) and NIR S:G lignin ratio (S:G). Family, site and family within site were the two conditions tested.

E. grandis HS (n = 349)					
Variable	df	SumSq	MeanSq	F-value	Pr(>F)
Diameter at breast height (DBH					
Site	1	0.00	0.00	0.00	1.00
Family	1	1.20	1.20	1.48	0.23
Site and family	1	2.47	0.82	1.01	0.39
Height					
Site	1	0.00	0.00	0.00	1.00
Family	1	6.00	6.00	7.52	6.42E-03
Site and family	1	2.07	0.69	0.86	0.46
Volume					
Site	1	0.00	0.00	0.00	1.00
Family	1	4.60	4.60	5.75	0.02
Site and family	1	2.58	0.86	1.08	0.36
Wood density					
Site	1	0.00	0.00	0.00	1.00
Family	1	25.13	25.13	35.38	6.72E-09
Site and family	1	5.68	1.89	2.67	0.05
Dissolving pulp yield (dPY)					
Site	1	0.00	0.00	0.00	1.00
Family	1	0.06	0.06	0.08	0.78
Site and family	1	2.15	0.72	0.90	0.44
S:G lignin ratio					
Site	1	0.00	0.00	0.00	1.00
Family	1	8.39	8.39	11.11	9.54E-04
Site and family	1	7.98	2.66	3.52	0.02

* Statistically significant at $P < 0.05$

Supplementary Table 2.12 One-way ANOVA for corrected phenotypic data for the *E. grandis* HS family across site. the following traits were analysed; diameter at breast height (DBH), height, volume, wood density, NIR dissolving pulp yield (dPY) and NIR S:G lignin ratio (S:G). Family was the condition being tested at a 0.05 significance threshold. Family was the condition being tested.

Trait	Variable	df	Sum Sq	Mean Sq	F-value	Pr (>F)
E. grandis site165 (n = 105)						
DBH	Family	1	0.00	0.00	0.00	0.97
Height	Family	1	0.11	0.11	0.13	0.72
Volume	Family	1	0.35	0.35	0.42	0.52
Density	Family	1	10.20	10.20	14.23	2.70E-04*
dPY	Family	1	0.02	0.02	0.03	0.87
S:G	Family	1	15.17	15.17	22.71	6.21E-06*
E. grandis HS site166 (n = 62)						
DBH	Family	1	1.90	1.90	2.47	0.12
Height	Family	1	2.69	2.69	3.56	0.06
Volume	Family	1	3.72	3.72	5.04	0.03*
Density	Family	1	0.00	0.00	0.00	0.95
dPY	Family	1	1.93	1.93	2.47	0.12
S:G	Family	1	0.49	0.49	0.60	0.44
E. grandis HS site167 (n = 99)						
DBH	Family	1	0.02	0.02	0.03	0.86
Height	Family	1	4.32	4.32	5.68	0.02*
Volume	Family	1	0.22	0.22	0.27	0.60
Density	Family	1	7.42	7.42	10.65	1.52E-03*
dPY	Family	1	0.17	0.17	0.22	0.64
S:G	Family	1	0.02	0.02	0.02	0.89
E. grandis HS site168 (n = 83)						
DBH	Family	1	1.74	1.74	2.13	0.15
Height	Family	1	0.95	0.95	1.15	0.29
Volume	Family	1	2.90	2.90	3.61	0.06
Density	Family	1	13.19	13.19	20.62	1.93E-05*
dPY	Family	1	0.08	0.08	0.10	0.75
S:G	Family	1	0.70	0.70	0.88	0.35

* Statistically significant at $P < 0.05$

Supplementary Table 2.13 Two-way ANOVA of corrected phenotypic data for the *E. urophylla* HS family. The following traits were analysed; diameter at breast height (DBH), height, volume, wood density, NIR dissolving pulp yield (dPY) and NIR S:G lignin ratio (S:G). Site, family and the interaction between family and site were tested.

E. urophylla HS (n = 367)					
Variable	df	SumSq	MeanSq	F-value	Pr(>F)
Diameter at breast height (DBH)					
Site	1	0.40	0.13	0.14	0.94
Family	1	20.00	3.33	3.53	2.09E-03*
Site and family	1	8.90	0.69	0.73	0.74
Height					
Site	1	0.30	0.10	0.10	0.96
Family	1	25.30	4.22	4.11	5.33E-04*
Site and family	1	12.20	0.94	0.91	0.54
Volume					
Site	1	1.00	0.34	0.37	0.78
Family	1	18.00	3.01	3.22	4.29E-03*
Site and family	1	7.20	0.55	0.59	0.86
Wood density					
Site	1	2.14	0.71	0.82	0.48
Family	1	13.20	2.20	2.54	0.02*
Site and family	1	9.64	0.74	0.86	0.60
Dissolving pulp yield					
Site	1	1.08	0.36	0.45	0.72
Family	1	17.84	2.97	3.72	1.34E-03*
Site and family	1	14.12	1.09	1.36	0.18
S:G lignin ratio					
Site	1	0.73	0.24	0.29	0.83
Family	1	6.70	1.12	1.32	0.25
Site and family	1	18.74	1.44	1.70	0.06

* Statistically significant at $P < 0.05$

Supplementary Table 2.14 One-way ANOVA of corrected data for the *E. urophylla* HS family. The following traits were analysed; diameter at breast height (DBH), height, volume, wood density, NIR dissolving pulp yield (dPY) and NIR S:G lignin ratio (S:G). The test was performed across different sites with family as the condition being tested at a 0.05 significance level.

Trait	Variable	df	SumSq	MeanSq	F-value	Pr(>F)
E. urophylla HS site 165 (n = 100)						
DBH	Family	1.00	6.80	1.36	1.47	0.21
Height	Family	1.00	5.59	1.12	1.19	0.32
Volume	Family	1.00	5.71	1.14	1.22	0.31
Density	Family	1.00	4.93	0.99	1.13	0.35
dPY	Family	1.00	20.39	4.08	5.76	1.12E-04*
S:G	Family	1.00	11.18	2.24	2.77	0.02*
E. urophylla HS site 166 (n = 67)						
DBH	Family	1.00	5.13	1.71	2.30	0.09
Height	Family	1.00	8.77	2.92	4.26	8.37E-03*
Volume	Family	1.00	5.08	1.69	2.28	0.09
Density	Family	1.00	6.44	2.15	3.11	0.03*
dPY	Family	1.00	3.19	1.06	1.49	0.23
S:G	Family	1.00	2.90	0.97	1.35	0.27
E. urophylla HS site 167 (n = 141)						
DBH	Family	1.00	10.29	1.72	1.55	0.17
Height	Family	1.00	17.93	2.99	2.25	0.04*
Volume	Family	1.00	8.08	1.35	1.26	0.28
Density	Family	1.00	3.65	0.61	0.60	0.73
dPY	Family	1.00	4.88	0.81	0.89	0.51
S:G	Family	1.00	4.98	0.83	0.83	0.55
E. urophylla HS site 168 (n = 59)						
DBH	Family	1.00	6.68	1.34	1.67	0.16
Height	Family	1.00	5.21	1.04	1.26	0.30
Volume	Family	1.00	6.37	1.27	1.58	0.18
Density	Family	1.00	7.83	1.57	2.29	0.06
dPY	Family	1.00	3.52	0.70	0.92	0.48
S:G	Family	1.00	6.38	1.28	1.80	0.13

* Statistically significant at $P < 0.05$

Chapter 3

Analysis of hybrid incompatibility in a *Eucalyptus* multi-parent mapping population

Julia Candotti[1], Marja M. O'Neill[1], S. Melissa Reynolds[1], Roobavathie Naidoo[2], Nicoletta Jones[2], Eshchar Mizrachi[1], Alexander A. Myburg[1*]

[1] *Department of Biochemistry, Genetics and Microbiology, Forestry and Agricultural Biotechnology Institute (FABI), University of Pretoria, Private bag X20, Pretoria 0028, South Africa*

[2] *Sappi Forests Research, Shaw Research Centre, PO Box 473, Howick, 3290, South Africa*

This research chapter has been prepared in the format required for submission to a peer-reviewed journal (*Tree Genetics and Genomes*). I performed all analyses in this manuscript and prepared the manuscript. Mrs M.M. O'Neill and Ms S.M. Reynolds provided technical support and advise on data analysis throughout the project. Ms R. Naidoo and Dr N. Jones constructed and maintained the multi-parent population, provided all sample tissue. Prof E. Mizrachi co-supervised the project. Prof A.A. Myburg conceived and supervised the project as well as provided valuable revisions for the manuscript.

3.1 Abstract

Understanding hybrid compatibility is important for efficient interspecific hybrid breeding programmes and for understanding natural speciation processes. Pre- and postzygotic reproductive barriers are two mechanisms affecting hybrid compatibility. Segregation distortion of DNA marker alleles, which is a deviation from expected Mendelian inheritance, can be used to identify regions of parental genomes underlying hybrid incompatibility. Multiparent populations, constructed by crossing a number of diverse founders, provide a resource for dissecting pre- and postzygotic barriers segregating in multiple parental genomes within half-sib (HS) and full-sib (FS) families. A *Eucalyptus* multi-parent mapping population was constructed by crossing nine *E. grandis* pollen parents and eight *E. urophylla* seed parents. This provided the opportunity to analyse the compatibility of parental alleles of one species across multiple families of a second species. In this study, we analysed segregation distortion patterns of SNP markers included in framework genetic linkage maps for one *E. grandis* pollen parent and one *E. urophylla* seed parent. Segregation distortion was analysed within HS families, FS families and in different environments (sites). The percentage of distorted SNP markers varied greatly between HS and FS families as well as within the different sites. Analysis of segregation patterns within sites suggested that environment-dependent interactions between parental genomes result in unique patterns of segregation distortion. We also used the segregation patterns of dead and living individuals of a single FS family to further dissect pre- and postzygotic reproductive barriers. This study showed that a large number of regions underlie hybrid compatibility of *E. grandis* and *E. urophylla*.

3.2 Introduction

According to Mendelian inheritance, alleles at heterozygous loci segregate in equal proportions from parents to progeny. Segregation distortion (SD) occurs when parental alleles deviate from the Mendelian expectation of equal segregation. Analysis of significantly distorted DNA markers in interspecific crosses can be used to obtain genome-wide evidence of pre- and postzygotic factors

contributing to hybrid incompatibility. Hybrid incompatibility is a type of natural reproductive barrier, but here we will discuss hybrid incompatibility from the perspective of plant hybrid breeding. Prezygotic barriers include habitat, flowering time, pollen germination and pollen tube formation which prevent fertilization (Rieseberg and Blackman 2010). Postzygotic barriers include differences in chromosome structure or genic interactions which result in hybrid sterility, necrosis or a reduction in fitness (Maheshwari and Barbash 2011).

Different evolution of genes as well as chromosomal rearrangements in species can cause barriers to hybrid compatibility (Burke and Arnold 2001). When two genes evolve independently in different species, they can cause negative genic interactions when combined in hybrids. This type of incompatibility follows the Dohzhansky-Muller model (Dobzhansky 1937; Muller 1942) which states that two loci interact in a negative manner and can result in hybrid necrosis. Chromosomal rearrangements can be due to insertions, deletions and inversions within the two species genomes. When the genomes are combined in a hybrid individual, the chromosomal rearrangements can affect meiosis in the hybrids resulting in hybrid sterility (Maheshwari and Barbash 2011). Therefore, for plant breeding, it is important to identify the causes and regions of the genome underlying hybrid incompatibility, to determine which parents can be crossed and will result in viable progeny.

A commonly used method for identifying hybrid incompatibility loci is the mapping of quantitative trait loci (QTL) underlying a hybrid incompatibility trait. Using this approach, QTL underlying seed viability in a F_1 interspecific cross in Arabidopsis were identified (Burkart-Waco *et al.* 2012). In the study, seed survival, of hybrid seed, was assessed to identify QTL underlying postzygotic hybrid incompatibility. A total of seven QTL were identified in the study, all of which had epistatic interactions detected. The authors concluded that there are multiple loci which interact and underlie hybrid seed survival in Arabidopsis. In another study, Yu *et al.* (2018), used this method to identify QTL underlying hybrid male sterility in rice. Using a F_2 backcross population of two diverged rice

species, they were able to identify four QTL underlying hybrid male sterility. This information was then used to identify the causal genes and the mechanisms underlying hybrid male sterility. While this study was successful in being able to identify QTL, QTL mapping can be limited due to the low resolution of most genetic linkage mapping studies, although experimental designs such as multi-parent populations can increase the power and resolution (Yu *et al.* 2008; Kover *et al.* 2009). Another limitation is that only traits of interest are analysed and not all QTL underlying barriers to hybrid incompatibility will be identified.

Another method to identify hybrid incompatibility loci is the analysis of segregation distortion. This allows for the identification of regions of the genome which may be harbouring genetic factors contributing to reproductive barriers. In a study by Li *et al.* (2019), nine F_1 inter-subspecific populations of rice were analysed for segregation distortion. A total of 61 significantly distorted regions were identified, of which 37 had previously been shown to underlie hybrid sterility. This study demonstrates the power that analysing segregation distortion has to identify a large number of genome-wide barriers to hybrid compatibility. Segregation distortion analysis can be used to identify both pre- and postzygotic barriers to hybrid compatibility. In a study by Bodénès *et al.* (2016), segregation distortion was used to identify possible prezygotic incompatibility loci in oak trees. The study made use of two intra- and two interspecific F_1 FS families to construct genetic linkage maps and analyse segregation distortion. A total of 79% of significantly distorted markers had a paternal origin which reflects that the segregation distortion of these markers are due to pollen incompatibilities. Therefore, the analysis of segregation distortion allows for the mechanisms underlying hybrid incompatibility to be identified.

Genotype-by-environment interaction is an important phenomenon in quantitative genetics and is known to affect many plant traits. However, most studies which aim to dissect hybrid incompatibility have been performed on single sites and have not taken into account the effect of the environment on

reproductive barriers. In a study, Chen *et al.* (2014) identified a two-locus interaction causing a reproductive barrier which was affected by temperature. They analysed an F_1 interspecific population in rice and found that a negative interaction between two loci caused hybrid weakness. The interaction was only induced at high temperatures which shows that the incompatibility barrier was environment-dependent. Therefore, hybrid incompatibility studies should be performed across multiple environments to determine environment dependent interactions contributing to hybrid compatibility.

Nested association mapping (NAM) populations are constructed by crossing a number of diverse founders. These populations are advantageous for studying hybrid compatibility as a large number of F_1 hybrids can be generated from a small number of diverse parents and full-sib families are nested within half-sib families. Multi-parent mapping populations have been shown to exhibit segregation distortion (McMullen *et al.* 2009; Song *et al.* 2017). McMullen *et al.* (2009) found that 17% of markers in the maize intraspecific NAM population were significantly distorted at a 0.05 significance level. The authors also found that of the five most highly distorted regions, four could be explained by previously identified genetic factors. Song *et al.* (2017) analysed an intraspecific soybean NAM population for segregation distortion. They found that the segregation distortion varied between the families and that there was an average of 3.75% of significantly distorted loci at a 0.01 significance level. These studies show that, due to the population design, segregation distortion is present in multi-parent populations and can be analysed within HS and FS families. These studies were performed in intraspecific hybrids. We expect to see much more segregation distortion within multi-parent mapping populations of interspecific hybrids as interspecific hybrid populations have been shown to have higher amounts of segregation distortion (Kullan *et al.* 2012b; Bodénès *et al.* 2016)

Eucalyptus is commonly planted as F_1 interspecific hybrid clones in commercial plantations. One of the most common hybrid combinations is between *E. grandis* and *E. urophylla* (Bison *et al.* 2006). This is due to the good growth and wood properties of *E. grandis* and the higher amount of disease

resistance in *E. urophylla* (Retief and Stanger 2009). There have been studies which have analysed segregation distortion within *Eucalyptus* hybrids (Grattapaglia and Sederoff 1994; Myburg *et al.* 2004; Freeman *et al.* 2006; Kullan *et al.* 2012b). Myburg *et al.* (2004) analysed segregation distortion patterns in *E. grandis* and *E. globulus* F_2 interspecific hybrid families to identify postzygotic barriers to hybridisation. They found that 27.7% of the AFLP markers analysed, were significantly distorted at a 0.05 level of significance. From this the authors concluded that postzygotic barriers were underlying interspecific hybrid incompatibility. Limitations of previous studies in *Eucalyptus* to identify genetic factors underlying hybrid compatibility is that they were performed across single environments and only included a few parental genotypes. The development of a replicated *Eucalyptus* interspecific hybrid multi-parent population enables the analysis of genome-wide transmission ratio distortion patterns of parental alleles within F_1 hybrid progeny in HS families, FS families and in different replicated environments.

This study aimed to analyse segregation patterns of SNP markers included in framework genetic linkage maps of a *E. grandis* pollen parent and a *E. urophylla* seed parent used in a nine by eight nested F_1 hybrid crossing design. Due to the population design, we were able to analyse segregation patterns within HS families, FS families and different environments. We also used segregation distortion patterns in genotyped dead and living trees of an intersecting FS family to distinguish between pre- and postzygotic incompatibilities. We hypothesized that hybrid incompatibility will manifest as segregation distortion across the genome and that specific interactions between parental genomes as well as with environmental factors, will cause different patterns of segregation distortion for the FS families in different environments. Here we show that there is significant segregation distortion of varying patterns within the HS families, FS families and different environments and that there is likely a genotype-by-environment effect on hybrid compatibility factors. These results suggest that hybrid incompatibility involves complex genetic interactions as well as genotype-by-environment effects.

3.3 Materials and Methods

3.3.1 Plant material and SNP genotyping

This study focused on SNP markers included in framework genetic linkage maps of one *E. grandis* HS and one *E. urophylla* HS family of the *Eucalyptus* multi-parent mapping population described in Chapter 2. Briefly, leaf or wood (cambium) tissue were collected from different living trees between three months to four years of age and genomic DNA extracted. Samples were SNP genotyped using the *Eucalyptus* EUChip60K SNP chip (Silva-Junior *et al.* 2015), paternal or maternal informative markers were identified and framework genetic linkage maps constructed in JoinMap®41 (Van Ooijen 2006). We used SNP genotypes for five FS families within the *E. grandis* HS family and seven FS families in the *E. urophylla* HS family. The population was planted across four sites in South Africa as part of a commercial F_1 hybrid trial series (Supplementary Table 2.1, Supplementary Table 3.1, 3.2). The phenotype of the trees were recorded at four years of age, at which time trees were classified as dead or alive. We therefore had the genotypes of both dead and living trees as we had sampled living trees which died between sampling and phenotyping.

3.3.2 Segregation distortion analysis

SNP markers included in the framework genetic linkage maps from Chapter 2 were re-phased following genetic map construction. Individuals with a phase of 1 (output of JoinMap®4.1 (Van Ooijen 2006) had genotypes converted from nn to np and np to nn. Individuals were separated into the following categories; HS families, FS families, HS family per site, and the intersecting FS family was separated into dead and living trees (classified during phenotyping at four years). Segregation distortion was quantified for each category by: (np allele frequency – 0.5) x 100 (Myburg *et al.* 2004). A chi-square test was performed, at a 0.05 significance level, to compare the observed genotypic ratio of each SNP marker with that of the expected genotypic ratio.

3.4 Results

3.4.1 Segregation distortion within HS families

A large amount of variation was observed in seed set from the current set of crosses and a very similar pattern was observed in a repeat of a subset of crosses (results not shown, Sappi Forest Research, Kwambonambi, South Africa) suggested differential compatibility of the parental genotypes used for F_1 hybrid progeny trials. We therefore investigated whether such incompatibility manifested as segregation distortion in the genetic maps of the parents which may point to specific loci which could underlie such incompatibilities. We first investigated whether there was segregation distortion within each of the paternal *E. grandis* or maternal *E. urophylla* alleles segregating in the two HS families. We analysed a total of 388 and 422 SNP markers in 349 and 367 individuals for segregation distortion for the *E. grandis* and *E. urophylla* HS families, respectively (Supplementary File 3.1 and 3.2). We observed that 14.95% (*E. grandis* HS) and 29.38% (*E. urophylla* HS) of the SNP markers deviated significantly from what was expected under Mendelian segregation, at a 0.05 significance level (Table 3.1). These results represent the average segregation distortion in each HS family because each HS family consists of different FS families and were planted across multiple sites. We found that the segregation patterns differed between the *E. grandis* paternal and the *E. urophylla* maternal HS families. Linkage groups 6, 8, 9 and 10 in the *E. grandis* HS family had severely distorted regions (Figure 3.1), while linkage groups 3, 5, 6 and 10 had significant distortion in the *E. urophylla* HS family (Supplementary Figure 3.1). The regions of significant distortion on linkage group 6 and 10 in both HS families are in different genomic regions. These results suggest that there is a complex genetic basis underlying segregation distortion in this F_1 hybrid population.

3.4.2 Segregation distortion within FS families

Next we investigated whether the segregation distortion patterns differed between the FS families within each HS family regardless of site. We analysed SNP markers for segregation distortion in 27 to 104 individuals within each FS family across the four sites (Supplementary Table 3.1 and 3.2;

Supplementary File 3.1 and 3.2). We observed significant distortion (chi-square, $\alpha = 0.05$) between 3.09% and 20.10%, of the progeny in each FS family (Table 3.1). We found that the patterns of segregation distortion varied among FS families (Figure 3.1, Supplementary Figure 3.1). Additionally, we saw that in some FS families the segregation patterns of significantly distorted markers were in opposite directions (in some FS families, one allele was favoured, while in other families the alternative allele was favoured (e.g. blue blocks in Figure 3.1)). We found that these effects were often cancelled out in the HS family segregation pattern, which shows the importance of analysing each FS family separately. These results suggest that there are specific genetic interactions between the different parental genomes.

3.4.3 Segregation distortion within sites

Next, we assessed whether segregation distortion patterns were similar across the four sites for each HS family (Figure 3.2, Supplementary Figure 3.2, Supplementary File 3.3 and 3.4). The number of individuals analysed per site were between 59 to 141 and the percentage of significantly distorted markers, at a 0.05 significance level, ranged from 1.55% to 15.64% (Table 3.1). We observed that the patterns of distortion varied across the different sites in both the *E. grandis* and *E. urophylla* HS families, and no sites had the same patterns of distortion across the entire genome. These results suggest that each site has a different effect on the interaction between the parental genomes. However, not every FS family was planted on each site so we could not make inferences for all families. Additionally, we do see some regions (e.g. Chromosome 6, Figure 3.2.) which has significantly distorted loci across all sites and in the same direction which suggests that this region is not affected by the environment and may be of large effect.

3.4.4 Segregation distortion across site and FS family

Next, we wanted to investigate possible genotype-by-environment effects by analysing the segregation distortion for each FS family within every site. We asked whether the segregation

distortion patterns were similar for FS families on the same site as well as for individual FS families planted across multiple sites. The number of individuals per individual FS family on one site ranged between 15 to 30 (minimum number of individuals was set at 15 for these analyses, Supplementary Table 3.1 and 3.2). We observed between 0 – 16.24% of the SNP markers to be significantly distorted at a 0.05 significance level in single FS families per site (Table 3.1, Figure 3.3, Supplementary Figure 3.3, 3.4, 3.5, 3.6, Supplementary Files 3.5 and 3.6). The segregation patterns varied greatly for a single FS family across multiple sites and different FS families on the same site. We also observed genomic regions which had opposite alleles favoured between FS families or sites such as Chr5 for site165 with multiple families and Chr6 for FS family FK593 across two sites (Figure 3.3). Together these results show that there are genetic interactions as well as environment dependent interactions (all of which should be postzygotic interactions) between the two genomes of different seed and pollen parents,.

3.4.5 Dissemination of pre- and postzygotic incompatibilities

Next we wanted to determine whether we could disseminate pre- and postzygotic factors contributing to hybrid compatibility. To do this, we analysed the segregation distortion of the parental alleles in 26 dead and 67 living trees of the intersecting FS family (FS family FK602 for paternal alleles, and FS family FK595 for maternal alleles. This is one FS family numbered differently for the maternal and paternal alleles, Supplementary File 3.7), which shares both of the parents analysed in this study (Figure 3.4). The rationale behind this is that the comparison of patterns of segregation distortion between dead and living trees, can help to identify potential pre- and postzygotic incompatibility factors. We observed that in some regions, the segregation distortion was in the same direction for dead and living trees (e.g. Chr3 of *E. urophylla* seed parent, Figure 3.4). We hypothesize that these regions could underlie a prezygotic barrier because the same allele was favoured in dead and living trees, suggesting that this allele passed through a prezygotic barrier. DNA was extracted from living trees of between 3 months to 4 years of age and trees were classified as dead or living upon

phenotyping at four years. Therefore postzygotic, pre-sampling selection is another possible explanation when we observe this pattern of segregation distortion between the dead and living trees. We also identified regions which showed segregation distortion in the opposite direction between dead and living trees (e.g. Chr1 of *E. urophylla* seed parent, Figure 3.4). We hypothesize that a postzygotic incompatibility barrier could underlie these regions, because living trees carrying one allele seemed to have a fitness advantage, while those carrying the alternative allele had a poor chance of survival. Due to both alleles being present in the plants at the time of sampling, a prezygotic factor could be ruled out, therefore the segregation distortion was likely due to a postzygotic factor.

From Figure 3.4, it can be seen that pre- and postzygotic barriers possibly operate in the parents and progeny. These results suggests that there may be many complex pre- and postzygotic interactions between parental genotypes which affect hybrid compatibility and viability. It is important to note that the intersecting family was planted across different sites, but site was not considered in this section due to small sample sizes. Site should not affect prezygotic incompatibilities, as the population was created through controlled crosses in a single site (nursery). However, site will most likely affect postzygotic incompatibilities as site was shown to affect segregation distortion in this study. We also expected that a genomic region underlying a prezygotic incompatibility would show the same segregation distortion pattern in the same FS family planted across multiple sites as environment should not affect prezygotic selection in this study. However, we do not observe any such pattern, therefore the results suggest mostly postzygotic, pre-sampling barriers as a more likely explanation when a ‘prezygotic’ pattern of segregation distortion is observed.

3.5 Discussion

Many studies have shown the value of multi-parent mapping populations for genetic dissection of complex traits. This study showed that multi-parent mapping populations can also be used for the analysis of reproductive barriers affecting hybrid compatibility. The population design of the

Eucalyptus multi-parent population allowed for the analysis of segregation distortion within HS families, FS families and different environments (sites). Numerous segregation distortion regions were identified with varying patterns across FS families, sites and FS families within sites. This suggests that hybrid compatibility involves many complex genetic interactions between parental genomic variation and environment.

This study was limited by the small sample size of FS families within sites. While the sample size was sufficient when analysing the segregation patterns across FS families, or across sites, these analyses did not take into consideration the genotype-by-environment effect. Therefore, the segregation distortion of single FS families within each site were analysed to determine the interaction between the parental genotypes and the environment. These results represent a more complete picture of how segregation distortion is affected by both the genetic interaction between the parental genotypes as well as the interaction with the environment.

Nevertheless, for these analyses, the sample sizes were small (between 15 – 30 individuals). When the sample size is small for a FS family on a site, there is the possibility that by chance, more individuals of one genotype were planted on a site than the alternative genotype which will result in false-positives for segregation distortion. The magnitude of the segregation distortion may also be inflated when the sample size is small, which further results in false-positive segregation distortion. Additionally, we may not have the statistical power to detect significant segregation distortion due to the small sample size. Therefore, while the results of this study can provide an insight into the way in which parental genomes interact with each other and with the environment, further studies with larger sample sizes are required to gain a more accurate understanding and fine-scale detection of the genetic basis underlying hybrid compatibility.

In this study, each marker was treated as independent, however, due to linkage, markers are not independent. Therefore, one could argue that the p-value should have been adjusted to account for multiple testing. Identifying regions of the genome that are in linkage and segregate together, was challenging in this population due to the large amount of variability present. A previous study by Myburg *et al.* (2004), used the assumption that each chromosome consists of two independent chromosome arms resulting in approximately 22 independent genomic regions in *Eucalyptus* (n = 11). The p-value was determined at a genome-wide threshold of 0.05/22 = 0.00227. However, even within each independent chromosome arm, there may be regions which segregate independently from each other. Previous studies on segregation distortion have used an uncorrected p-value of 0.05. Therefore, due to the challenges of identifying independent chromosomal regions, and in order to compare the results with previous studies, a p-value of 0.05 was used in this study. We are however, aware that this p-value may not be stringent enough and may result in many false-positive segregation distortion observations.

It is also important to note that epistatic interactions between loci were not analysed in this study. In *Eucalyptus*, Myburg *et al.* (2004) identified epistatic interactions in *E. grandis* and *E. globulus* backcross populations. Therefore, we expect that there will be epistatic interactions which will affect hybrid compatibility as it has been shown to play a role in *Eucalyptus* and other plant species (Rieseberg *et al.* 1996; Myburg *et al.* 2004). Therefore, future studies on segregation distortion in this population will need to analyse epistasis to improve our understanding of hybrid compatibility.

3.5.1 Identification of segregation distortion

We found that the percentage of segregation distortion within the HS families (14.95% *E. grandis* and 29.38% *E. urophylla*) of this study were slightly lower than previous segregation distortion studies which reported segregation distortion ranging from 27.5% to 36.3% (Myburg *et al.* 2004; Kullan *et al.* 2012b; Bartholomé *et al.* 2015). This could be due to this study consisting of a number

of FS families within each HS family and the population being planted across sites, whereas the previous studies have focused on single bi-parental crosses on a single site. When we analysed the segregation patterns within FS families and sites, we saw that there are regions in which one allele was favoured in one FS family/site while the alternative allele was favoured in another FS family/site. This resulted in the direction of segregation distortion averaging out in the HS families resulting in fewer significantly distorted markers. We observed a large variation in the percentage of significantly distorted markers of individual FS families on a single site. These results suggest interaction of genetic variation in parental genomes and environment.

A comparison of the results of this study with previous studies was limited due to different population designs and type of markers used. Despite this, we compared our results with that of the most recent genetic linkage maps for *E. grandis* and *E. urophylla* in which segregation distortion was analysed in a single, large FS family (Bartholomé *et al.* 2015). This study found that chromosomes 5, 6, 7 and 11 for *E. urophylla*, and chromosomes 1 and 3 for *E. grandis* had significantly distorted regions. In the current study, we found that for the *E. urophylla* HS family, significant distortion was seen on chromosomes 3, 5, 6 and 10, while significant distortion on chromosomes 7 and 11 were only seen when analysing individual FS families. No significant distortion was seen on chromosome 1 or 3 for the *E. grandis* HS family, but there was significant distortion on these chromosomes in some FS families. Regions which show similar patterns of segregation distortion across multiple studies and FS families, could indicate regions underlying common incompatibility loci, while regions which are unique to each FS family could indicate specific genetic interactions between parental genomes. The comparison of these studies again suggest that segregation distortion patterns vary between crosses of different parental genotypes suggesting that hybrid compatibility factors segregate in both parental species.

3.5.2 Causes of segregation distortion

Segregation distortion of alleles can occur due to pre- and postzygotic mechanisms as well as genetic load and meiotic drive. Genetic load occurs when there is an accumulation of deleterious alleles which can result in segregation distortion. In order to determine whether genetic load is the cause of segregation distortion, both intra and interspecific hybrids need to be analysed which was outside of the scope of this study (Bodénès *et al.* 2016). Meiotic drive is the when one allele is favoured over another allele during meiosis which will cause segregation distortion at the locus. In order to determine if segregation distortion is due to meiotic drive, the fertility of hybrids need to be analysed (Fishman and Willis 2005), which again was outside the scope of this study. In the current study, while we refer to hybrid incompatibilities, it is important to note that prezygotic mechanisms result in hybrid incompatibility while postzygotic mechanisms are more likely to result in hybrid viability. Causes of prezygotic hybrid incompatibility which can result in segregation distortion include pollen tube growth rate (Rieseberg and Blackman 2010), meiotic drive (Cameron and Moav 1956) and gametophytic incompatibility (Lin *et al.* 1992). Through the analysis of patterns of segregation distortion of dead and living trees of the intersecting family, we were able to identify regions underlying potential prezygotic hybrid incompatibility barriers.

It is important to note that we cannot directly distinguish between prezygotic and postzygotic but pre-sampling incompatibilities in this study. This is because we do not have the allele frequencies before and directly after fertilisation. We also do not have significant information regarding the survival of the trees (and age of death) prior to sampling to determine how much of the distortion could be due to postzygotic, pre-sampling incompatibilities. Sampling of different trees was performed at various life-stages of the trees (between 3 months to four years), which limits our ability to determine at which stage the postzygotic, pre-sampling barrier could have occurred. However, we would expect that genomic regions underlying prezygotic compatibility barriers would be seen across individual FS families on different sites, as the environment in which the progeny is planted would not affect

prezygotic barriers. In this study, we did not see any evidence of this. This could be due to the small sample size which limited our ability to detect segregation distortion. Additionally, it is possible that some chromosomal regions are affected by overlapping pre- and postzygotic factors (where one allele is favoured in the prezygotic stage and the alternative in favoured in the postzygotic stage). This could result in the masking of either the pre- or postzygotic factor. Therefore, future studies focused on identifying segregation distortion at different life stages can help to identify the causes of hybrid incompatibility.

Postzygotic incompatibility can include genetic incompatibility and chromosome structural variation (this usually has more of an effect on F_2 progeny) between the parental genomes. In this study, it was more likely that the cause of postzygotic barrier was due to genetic incompatibility between the parental genomes rather than structural differences. Although we could not compare the order of markers between the two parental maps due to different markers being used (by default), comparison of the marker order with the *Eucalyptus grandis* v2 genome assembly (https://phytozome.jgi.doe.gov/) showed that there was high collinearity. There were some regions which showed marker order changes or markers which mapped to different linkage groups to what was expected. However, when these markers were found to be significantly distorted, the markers surrounding them were also found to be significantly distorted with higher chi-square values. This suggests that chromosomal rearrangements potentially only have a small, if any, effect on segregation distortion. These results are in agreement with a previous which did not find any major chromosomal rearrangements between *E. grandis* and *E. globulus* and which support genic incompatibilities rather than structural incompatibilities (Hudson *et al.* 2012; Myburg *et al.* 2014).

Genic incompatibility between the parental genomes is a mechanism that can result in postzygotic barriers. This type of incompatibility often follows the Dobzhansky-Muller (DM) diverged genes model which states that loci which are compatible in the ancestral state, diverge independently during

evolution, resulting in a negative interactions when combined in interspecific hybrids (Dobzhansky 1937; Muller 1942). This could result in segregation distortion as only some genetic combinations are compatible. In the current study, we cannot determine negative genic interactions between multiple loci of the parental genomes (epistasis) by looking at the segregation distortion patterns alone and we did not analyse epistatic interactions. However, this type of incompatibility has previously been shown in *E. grandis* and *E. globulus* (Myburg *et al.* 2004) and we expect it will also be present within this population.

Genotype-by-environment interactions observed in this study suggests postzygotic barriers because the population was constructed by controlled pollination on a single site. The seeds were germinated and planted in the same nursery before planting on four different sites. Therefore, we hypothesize that observed differences in segregation patterns of individual FS families across multiple sites are likely due to postzygotic factors. These post-zygotic factors are potentially influenced by the environment (seen by the different patterns of segregation distortion across sites) and result in differences in hybrid viability across the sites. For example, analysis of the FS family FK593 across two sites show that in one region of significant distortion on Chr6, one allele is favoured in one site, while the alternative allele is favoured on the other site. This suggests that the environment is affecting the segregation distortion at this region. We do not see many examples of this, which again could be due to the small sample size. Therefore, future studies with larger samples sizes are required to determine the effect of environment on hybrid compatibility.

3.5.3 Application for industry

Understanding the genetic basis of hybrid compatibility is important for the forestry industry as it will allow the design of more efficient hybrid breeding programs. Artificial hybridisation between species which do not naturally occur together is important for breeding programs as it increases the genetic diversity and allows for combining of favourable traits into a common genetic background.

However, many interspecific crosses in *Eucalyptus* are not successful (Griffin *et al.* 1988). Therefore, if we could improve our understanding of hybrid incompatibility in *Eucalyptus* it would allow us to make more informed decisions about which species or even individual trees to cross.

From the *Eucalyptus* multi-parent population, we can see that there is a large amount of variability in the success of the crosses between different *E. grandis* pollen parents and *E. urophylla* seed parents. The *Eucalyptus* multi-parent population contains a number of diverse parents, however, due to hybrid incompatibility, a large proportion of this diversity is likely lost in the F_1 progeny. Therefore, the ability to determine which parental genotypes are compatible and will result in successful progeny in specific or a wide variety of environments is important for monitoring genetic diversity and improving the efficiency of hybrid breeding programmes. Additionally, it will be important to identify regions underlying hybrid incompatibility factors which are fixed versus segregating in the population. In this study, the factors identified are segregating due to the nature of informative markers analysed (heterozygous markers). We also selected the two HS families which yielded the highest number of progeny. This suggests that the crosses analysed in this study were the most successful and had the least lethal incompatibility factors when compared to the rest of the population. Therefore, we can use the segregating loci to identify combinations which are successful and yield viable progeny.

3.6 Conclusion and future prospects

This study was able to use genome-wide SNP linkage maps and analysis of segregation distortion to identify regions of parental genomes which possibly underlie hybrid incompatibility. The ability to analyse segregation patterns across multiple FS families and environments provided an insight into how genetic interactions between parental genomes may differ and how the environment can affect these interactions. We were also able to identify potential pre- and postzygotic incompatibilities, based on segregation distortion patterns of dead and living trees. This study suggests that hybrid

incompatibility is affected by complex genetic interactions, and that multi-parent populations can allow dissection and identification of compatible genotypes. However, larger sample sizes are required in order to improve our understanding of hybrid compatibility.

3.7 Acknowledgments

This work was funded by Sappi Forest research (Hilton, KZN, South Africa), the Department of Science and Technology (DST), Technology Innovation Fund (TIA) and the National Research Foundation (NRF). JC acknowledges the NRF for MSc bursary support.

3.8 Tables

Table 3.1 Summary of significantly distorted markers for each FS family per site. Segregation distortion was analysed for FS families with more than 15 individuals on a site with a significance threshold of 0.05. The overall family and overall site values are the percentage of distorted markers within the entire FS family or across families within the entire site (regardless of the number of individuals for each FS family on a site) and is not simply the total of the column or row. The value in the cell of both the overall family and overall site is the percentage of distorted markers in the entire HS family.

	Distorted markers (%)				
Site	Site165	Site166	Site167	Site168	Overall Family
Family	*E. grandis* HS family				
FK599	5.67	-	12.11	-	5.67
FK600	0.52	6.19	2.06	4.38	3.09
FK601	14.95	6.44	1.29	9.79	20.10
FK602	8.76	5.15	3.09	0.00	7.99
FK603	16.24	-	-	-	8.51
Overall Site/HS	4.65	8.51	7.47	1.55	14.95
Family	*E. urophylla* HS family				
FK592	4.03	-	5.45	-	9.48
FK593	8.53	-	6.87	-	11.85
FK594	5.69	2.37	-	-	6.64
FK595	5.45	4.27	11.14	7.58	10.43
FK596	-	-	13.03	-	4.03
FK604	-	-	3.08	-	5.45
FK605	-	-	15.88	-	9.48
Overall Site/HS	11.09	11.14	15.64	1.9	29.38

3.9 Figures

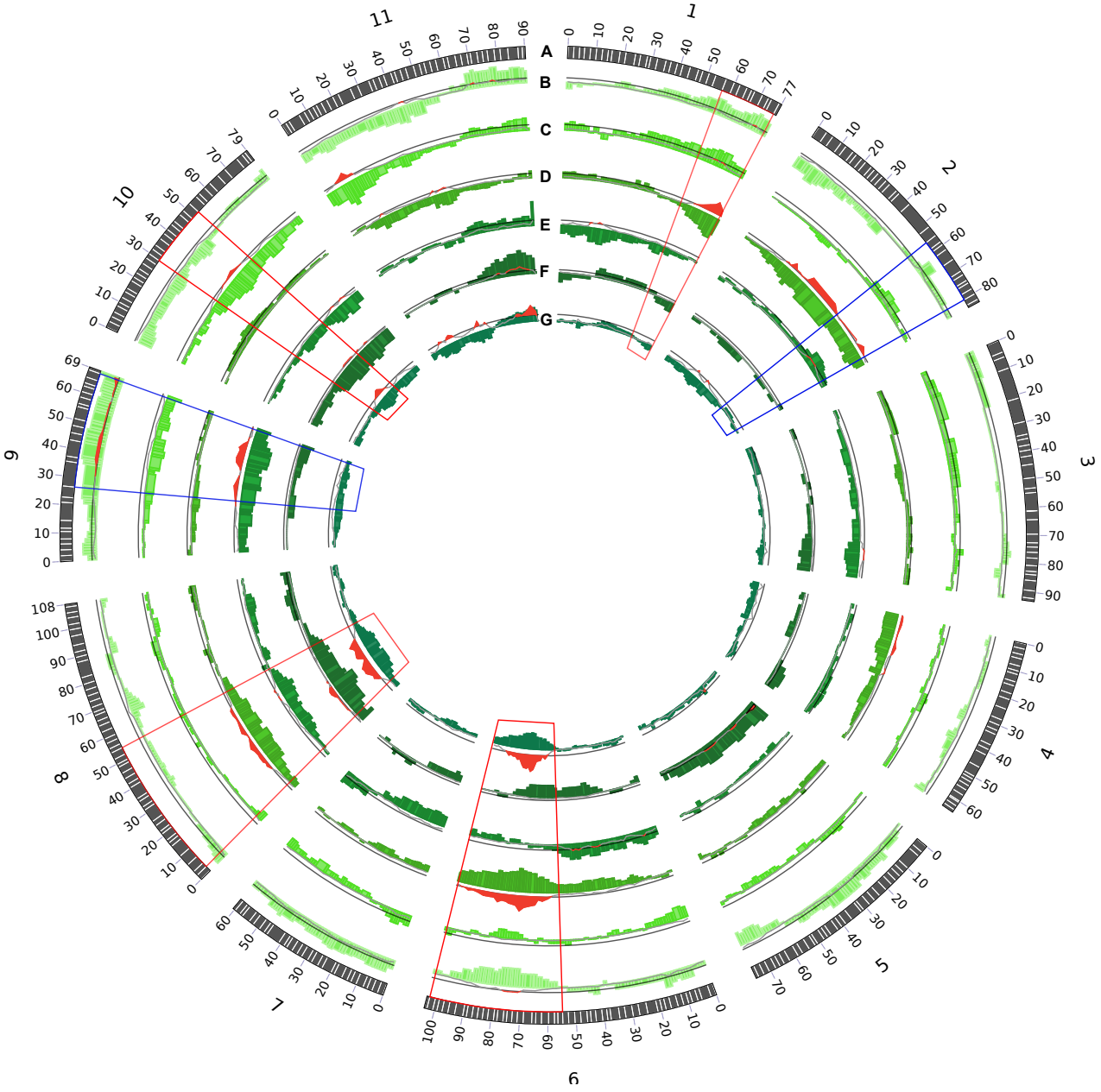


Figure 3.1 (Legend on page 140)

Figure 3.1 Segregation distortion patterns for each FS family within the *E. grandis* HS family. A. The outer track in black represents the 11 chromosomes and the white lines represent the markers included in the framework genetic linkage map with distance in cM. **B-F** represent the segregation distortion in each FS family (Supplementary File 3.1). Each green bar represents a region surrounding a marker and shows the direction and percentage deviation calculated by (np allele frequency – 0.5) x 100 (Myburg *et al.* 2004). The straight black line represents the chi-square critical value at a 0.05 significance level. The grey line with red segments represents the chi-square test statistic for deviation from the expected 1:1 segregation ratio for each marker. The red segments of the line and the shaded red regions show regions with significant distortion. **B.** FS family FK599. **C.** FS family FK600. **D.** FS family FK601. **E.** FS family FK602. **F.** FS family FK603. **G.** Entire *E. grandis* HS family. The blue rectangles show regions with opposite directions of segregation distortion in some FS families. Red rectangles represent regions which show the same direction of segregation distortion in all FS families.

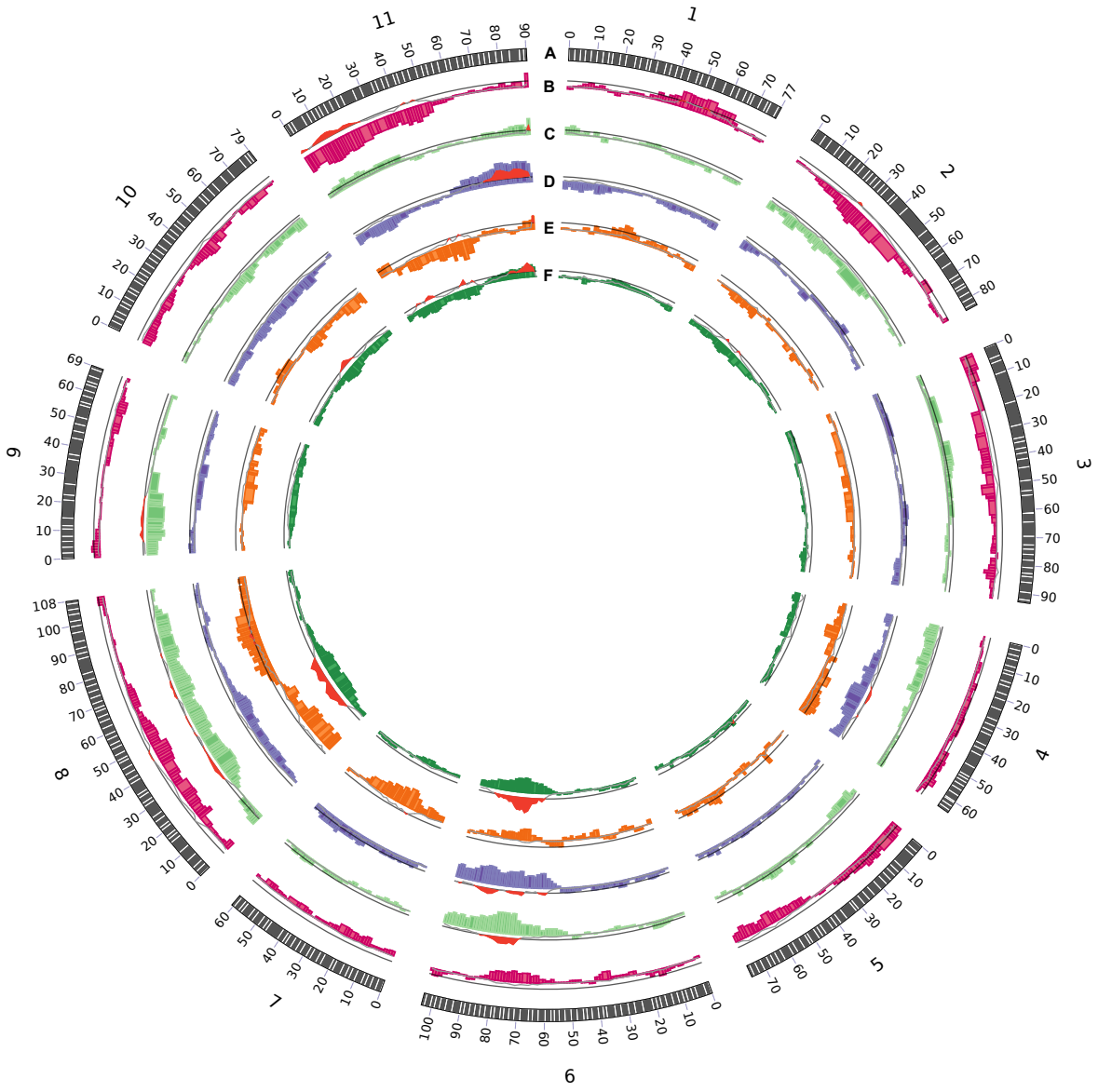


Figure 3.2 Segregation distortion patterns for the *E. grandis* HS across four sites. A. Black outer track represents the 11 linkage groups and the white lines represent the markers included in the framework genetic linkage map with distance in cM. **B-F** represent the segregation distortion patterns in each of the four sites and the entire HS family (Supplementary File 3.3). Each coloured bar represents a region surrounding a marker and shows the direction and percentage distortion of the marker calculated by (allele frequency – 0.5) x 100 (Myburg *et al.* 2004). The straight grey line represents the chi-square critical value at a 0.05 significance level. The grey line with red segments represent the test statistic for deviation from the expected 1:1 segregation distortion for each marker, with the red regions representing significant distortion. **B.** Site165. **C.** Site166. **D.** Site167. **E.** Site168. **F.** Entire *E. grandis* HS family across sites.

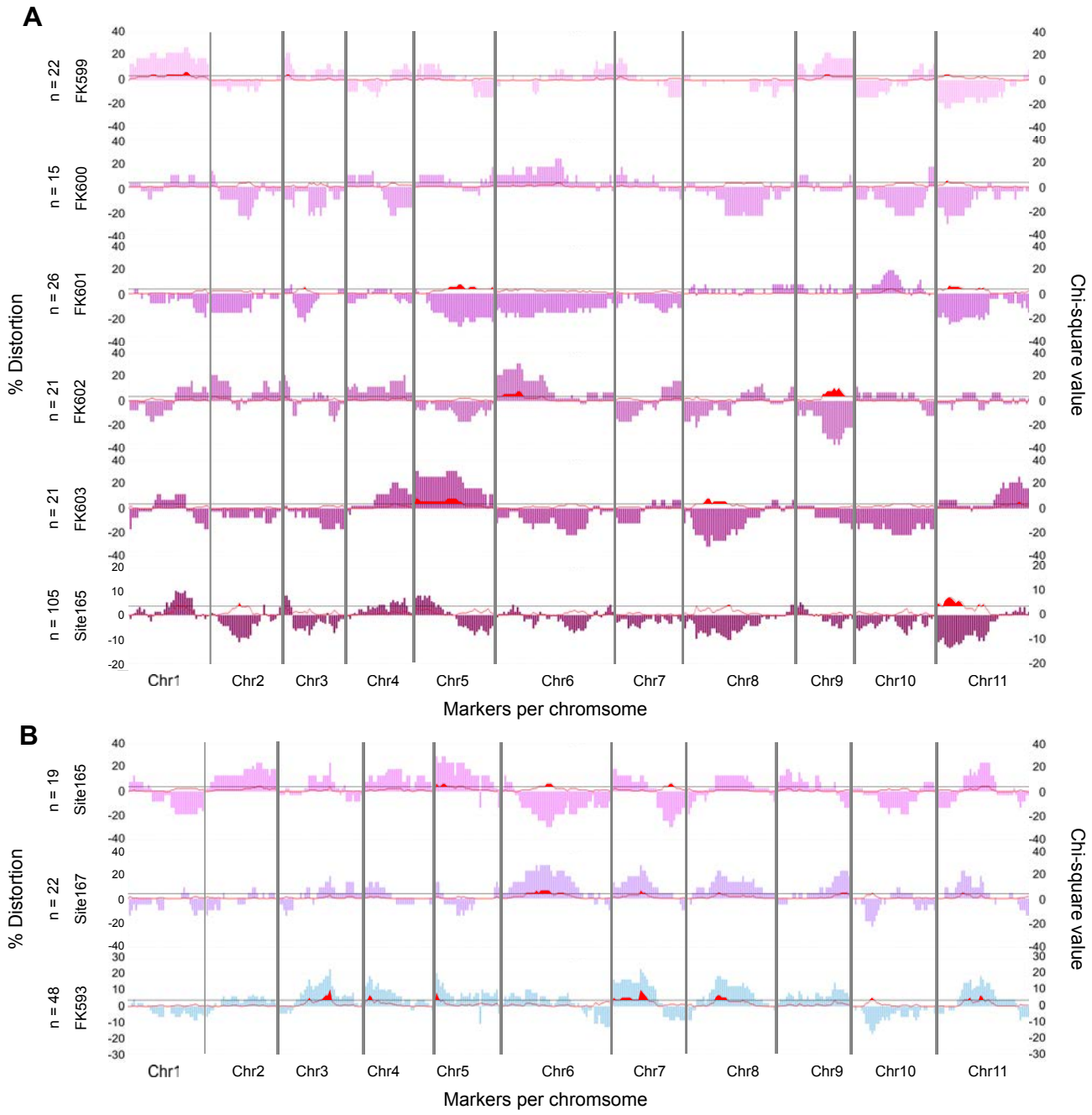


Figure 3.3 (Legend on page 143)

Figure 3.3 Segregation distortion magnitude, direction and distribution of a single site with multiple FS families and a single FS family across multiple sites. Each vertical bar represents the direction and percentage deviation calculated by (allele frequency – 0.5) x 100 (Myburg et al 2004, primary y-axis). The black horizontal line represents the chi-square critical value of 3.841 and the red line represents the chi-square test statistic for deviation from the expected 1:1 segregation distortion for each marker (secondary y-axis). **A.** Segregation distortion of one site with five FS families for the *E. grandis* HS family (Supplementary File 3.5). In some regions (e.g. Chromosome 5), one allele is favoured in FS family FK601, while the alternative allele is favoured in FS family FK603. **B.** Segregation distortion of a single *E. urophylla* FS family across two sites (Supplementary File 3.6). In some regions (e.g. Chromosome 6), one allele is favoured in site165 while the alternative allele is favoured in site167.

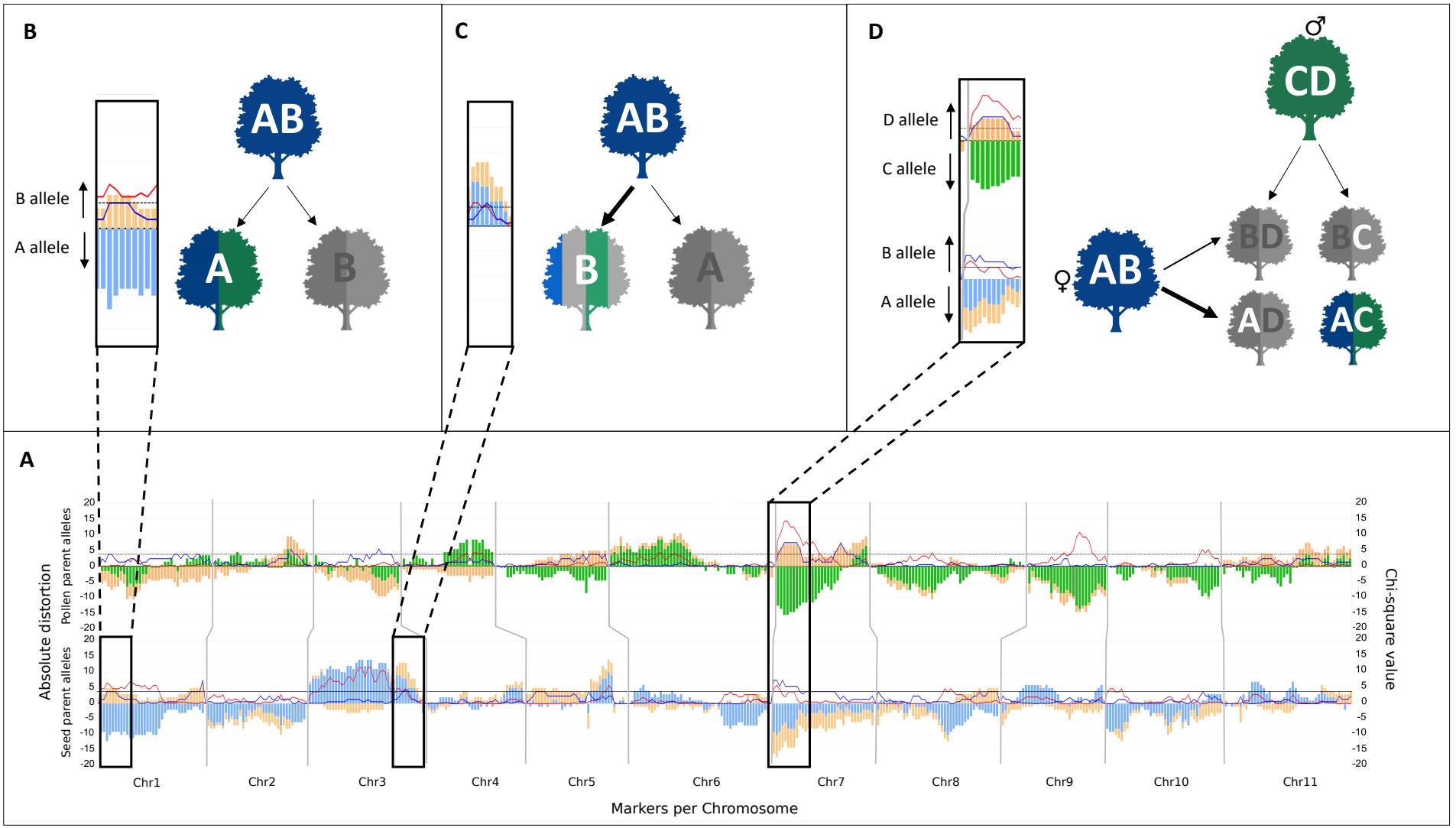


Figure 3.4 (Legend on page 145)

Figure 3.4 Identification of regions underlying pre- and postzygotic incompatibilities based on segregation distortion patterns in dead and alive trees. A. Segregation distortion patterns of SNP markers in dead (orange) and living trees (green for *E. grandis* and blue for *E. urophylla*) of the *E. grandis* pollen alleles (top pane) and *E. urophylla* seed parent alleles (bottom pane) for the intersecting FS family (FS family FK602 for the *E. grandis* pollen alleles and FS family FK595 for the *E. urophylla* seed parent alleles. This is one FS family, numbered differently based on which alleles are analysed, Supplementary File 3.7). Each bar represents a SNP marker and is ordered by chromosome (x-axis). The primary y-axis represents the absolute distortion (i.e. the number of individuals more than expected carrying an allele). The chi-square value of each marker for the living trees is represented by the red line and by the blue line for dead trees with the chi-square value represented on the secondary y-axis. The critical value (3.841) for the chi-square test at a 0.05 significance is represented by the black dotted horizontal line. **B.** Putative postzygotic barrier acting on the seed parent alleles. The segregation pattern of the dead trees (orange) show that trees carrying one allele (B allele) were more likely to die while the segregation patterns of the living trees (blue) show that trees carrying the alternative allele (A allele) were more likely to be alive. This suggests a postzygotic barrier. As is illustrated in the diagram where the seed parent tree (blue) carries the alleles A and B which are equally transmitted to the progeny. In the hybrid progeny (two-colored trees), trees carrying the A allele (tree with green and blue) are more likely to survive, while the trees carrying the B allele (grey trees) are more likely to die. **C.** Putative prezygotic (or postzygotic, pre-sampling) barrier on the seed parent alleles. The segregation distortion pattern show that the same allele (B allele) is more common in dead and living trees. This suggests that there was possibly unequal transmission of the seed parent alleles, which could be due to prezygotic or postzygotic, pre-sampling barriers that caused the B allele to be favored. This is visualized in the diagram where the seed parent (blue tree) is heterozygous (carrying A and B allele) but the trees carrying the B allele is seen more often in the progeny (blue, green and grey tree) when compared with trees carrying the A allele. **D.** Possible combination of pre- and postzygotic barriers of the pollen and seed parent alleles and progeny can be used to identify combinations of alleles which preferentially combine and survive. The segregation distortion of the pollen parent alleles (C and D alleles) suggests a post-zygotic barrier affecting the pollen alleles, resulting in trees carrying the A alleles surviving more often than those carrying the D allele. The segregation patterns of the seed parent alleles (A and B alleles) suggests prezygotic or pre-sampling barrier affecting these alleles with the A allele preferentially passing through the barrier. The diagram represents the result of the combination of the seed parent (blue tree) and pollen parent (green tree) alleles in the progeny. The pollen alleles segregate equally within the progeny (black arrows), trees carrying the D allele (grey trees) are more likely to die resulting in trees carrying the AC allele combination more likely to survive (green and blue tree and black text). The seed parent alleles do not segregate equally within the progeny (grey arrows), resulting in more trees carrying the A allele (green and blue tree and black text) than the B allele. Therefore, the most common allele combination in the progeny is AC.

3.10 References

Bartholomé J, Mandrou E, Mabiala A, Jenkins J, Nabihoudine I, Klopp C, Schmutz J, Plomion C, Gion J. 2015. High-resolution genetic maps of *Eucalyptus* improve *Eucalyptus grandis* genome assembly. *New Phytologist* 206: 1283–1296.

Bison O, Ramalho MAP, Rezende GDSP, Aguiar AM, de Resende MDV. 2006. Comparison between open pollinated progenies and hybrids performance in *Eucalyptus grandis* and *Eucalyptus urophylla*. *Silvae Genetica* 55: 192–196.

Bodénès C, Chancerel E, Ehrenmann F, Kremer A, Plomion C. 2016. High-density linkage mapping and distribution of segregation distortion regions in the oak genome. *DNA Research* 23: 115–124.

Burkart-Waco D, Josefsson C, Dilkes B, Kozloff N, Torjek O, Meyer R, Altmann T, Comai L. 2012. Hybrid incompatibility in Arabidopsis is determined by a multiple-locus genetic network. *Plant Physiology* 158: 801–812.

Burke JM, Arnold ML. 2001. Genetics and the fitness of hybrids. *Annual Review of Genetics* 35: 31–52.

Cameron DR, Moav R. 1956. Inheritance in *Nicotiana tabacum* XXVII. pollen killer, an alien genetic locus inducing abortion of microspores not carrying it. *Genetics* 42: 326–335.

Chen C, Chen H, Lin YS, Shen JB, Shan JX, Qi P, Shi M, Zhu MZ, Huang XH, Feng Q, *et al.* 2014. A two-locus interaction causes interspecific hybrid weakness in rice. *Nature Communications* 5: 3357.

Dobzhansky T. 1937. *Genetics and the origin of species*. New York: Columbia University Press.

Fishman L. and Willis JH. 2005. A novel meiotic drive locus almost completely distorts segregation in Mimulus (Monkeyflower) hyrbids. *Genetics* 169: 347-353

Freeman JS, Potts BM, Shepherd M, Vallancourt RE. 2006. Parental and consensus linkage maps of *Eucalyptus globulus* using AFLP and microsatellite markers. *Silvae Genetica* 55: 202–217.

Grattapaglia D, Sederoff R. 1994. Genetic linkage maps of *Eucalyptus grandis* and *Eucalyptus urophylla* using a pseudo-testcross: Mapping strategy and RAPD markers. *Genetics* 137: 1121–1137.

Griffin AR, Burgess IP, Wolf L. 1988. Patterns of natural and manipulated hybridisation in the genus *Eucalyptus* L'Herit - a review. *Australian Journal of Botany* 36: 41–66.

Hudson CJ, Freeman JS, Kullan ARK, Petroli CD, Sansaloni CP, Kilian A, Detering F, Grattapaglia D, Potts BM, Myburg AA, *et al.* 2012. A reference linkage map for *Eucalyptus*. *BMC Genomics* 13.

Kover PX, Valdar W, Trakalo J, Scarcelli N, Ehrenreich IM, Purugganan MD, Durrant C, Mott R. 2009. A multiparent advanced generation inter-cross to fine-map quantitative traits in *Arabidopsis thaliana*. *PLoS Genetics* 5: e1000551.

Kullan ARK, van Dyk MM, Jones N, Kanzler A, Bayley A, Myburg AA. 2012. High-density genetic linkage maps with over 2,400 sequence-anchored DArT markers for genetic dissection in an F2 pseudo-backcross of *Eucalyptus grandis* × *E. urophylla*. *Tree Genetics and Genomes* 8: 163–175.

Li G, Jin J, Zhou Y, Bai X, Mao D, Tan C, Wang G, Ouyang Y. 2019. Genome-wide dissection of segregation distortion using multiple inter-subspecific crosses in rice. *Science China Life Science* 62: 507–516.

Lin SY, Ikehashi H, Yanagihara S, Kawashima A. 1992. Segregation distortion via male gametes in hybrids between Indica and Japonica or wide-compatibility varieties of rice (*Oryza sativa L*). *Theoretical and Applied Genetics* 84: 812–818.

Maheshwari S, Barbash DA. 2011. The genetics of hybrid incompatibilities. *Annual Review of Genetics* 45: 331–355.

McMullen MD, Kresovich S, Villeda HS, Bradbury P, Li H, Sun Q, Flint-garcia S, Thornsberry J, Acharya C, Bottoms C, *et al.* 2009. Genetic properties of the maize nested association mapping population. *Science* 325: 737–740.

Muller HJ. 1942. Isolating mechanisms, evolution and temperature. *Biology Symposium* 6: 71–125.

Myburg AA, Grattapaglia D, Tuskan GA, Hellsten U, Hayes RD, Grimwood J, Jenkins J, Lindquist E, Tice H, Bauer D, *et al.* 2014. The genome of *Eucalyptus grandis*. *Nature* 510: 356–362.

Myburg AA, Vogl C, Griffin AR, Sederoff RR, Whetten RW. 2004. Genetics of postzygotic isolation in *Eucalyptus* : Whole-genome analysis of barriers to introgression in a wide interspecific cross of *Eucalyptus grandis* and *E . globulus. Genetics* 166: 1405–1418.

Van Ooijen JW. 2006. JoinMap ® 4, Software for the calculation of genetic linkage maps in experimental populations.

Retief ECL, Stanger TK. 2009. Genetic parameters of pure and hybrid populations of *Eucalyptus grandis* and *E. urophylla* and implications for hybrid breeding strategy. *Southern Forests: a Journal of Forest Science* 71: 133–140.

Rieseberg LH, Blackman BK. 2010. Speciation genes in plants. *Annals of Botany* 106: 439–455.

Rieseberg LH, Sinervo B, Linder CR, Ungerer MC, Ungerer MC, Arias DM. 1996. Role of gene interactions in hybrid speciation : Evidence from ancient and experimental hybrids. *Science* 272: 741–745.

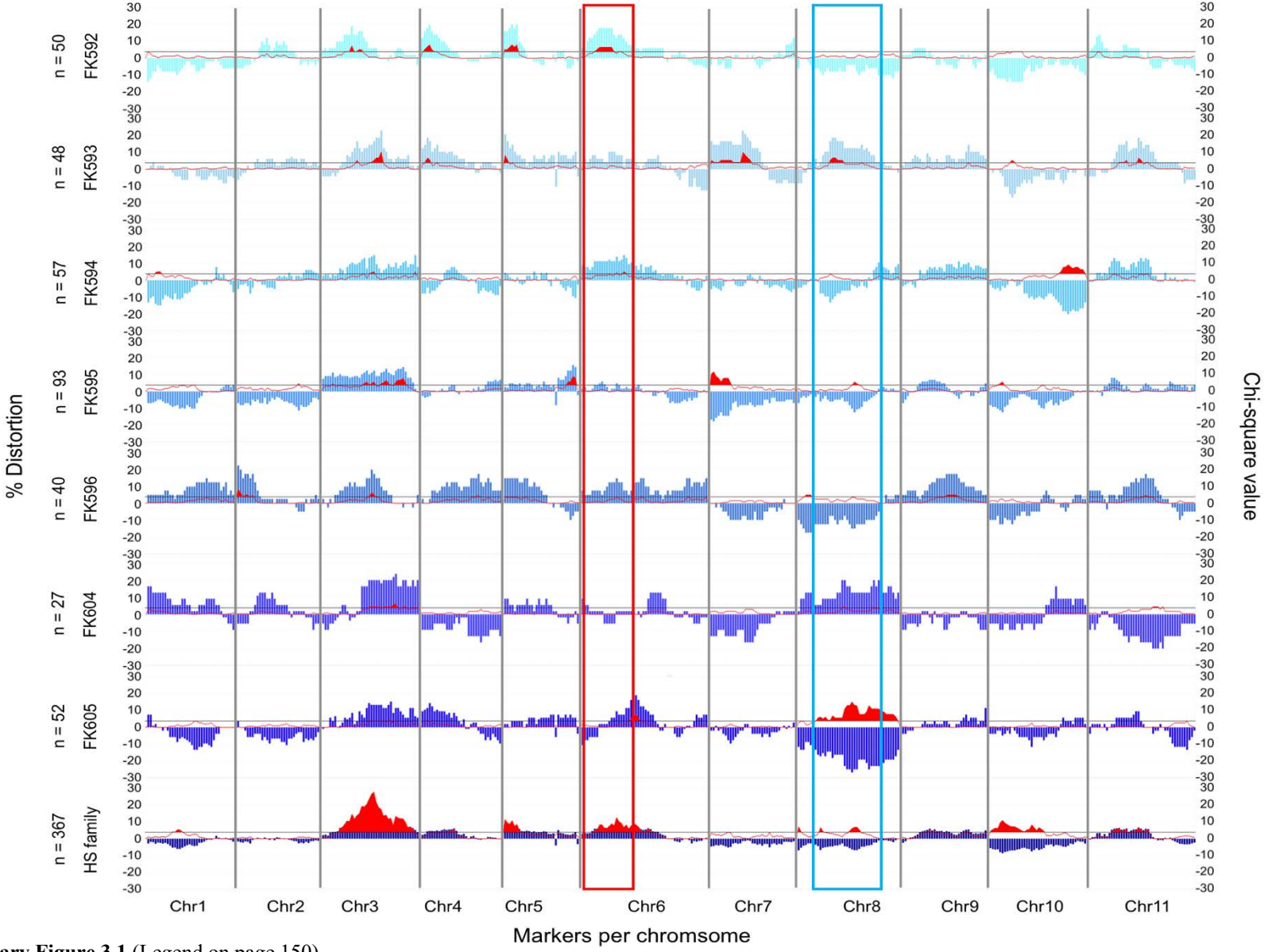
Silva-Junior OB, Faria DA, Grattapaglia D. 2015. A flexible multi-species genome-wide 60K SNP chip developed from pooled resequencing of 240 *Eucalyptus* tree genomes across 12 species. *New Phytologist* 206: 1527–1540.

Song Q, Yan L, Quigley C, Jordan BD, Fickus E, Schroeder S, Song B-H, Charles An Y-Q, Hyten D, Nelson R, *et al.* 2017. Genetic characterization of the soybean nested association mapping population. *The Plant Genome* 10: 2.

Yu J, Holland JB, McMullen MD, Buckler ES. 2008. Genetic design and statistical power of nested association mapping in maize. *Genetics* 178: 539–551.

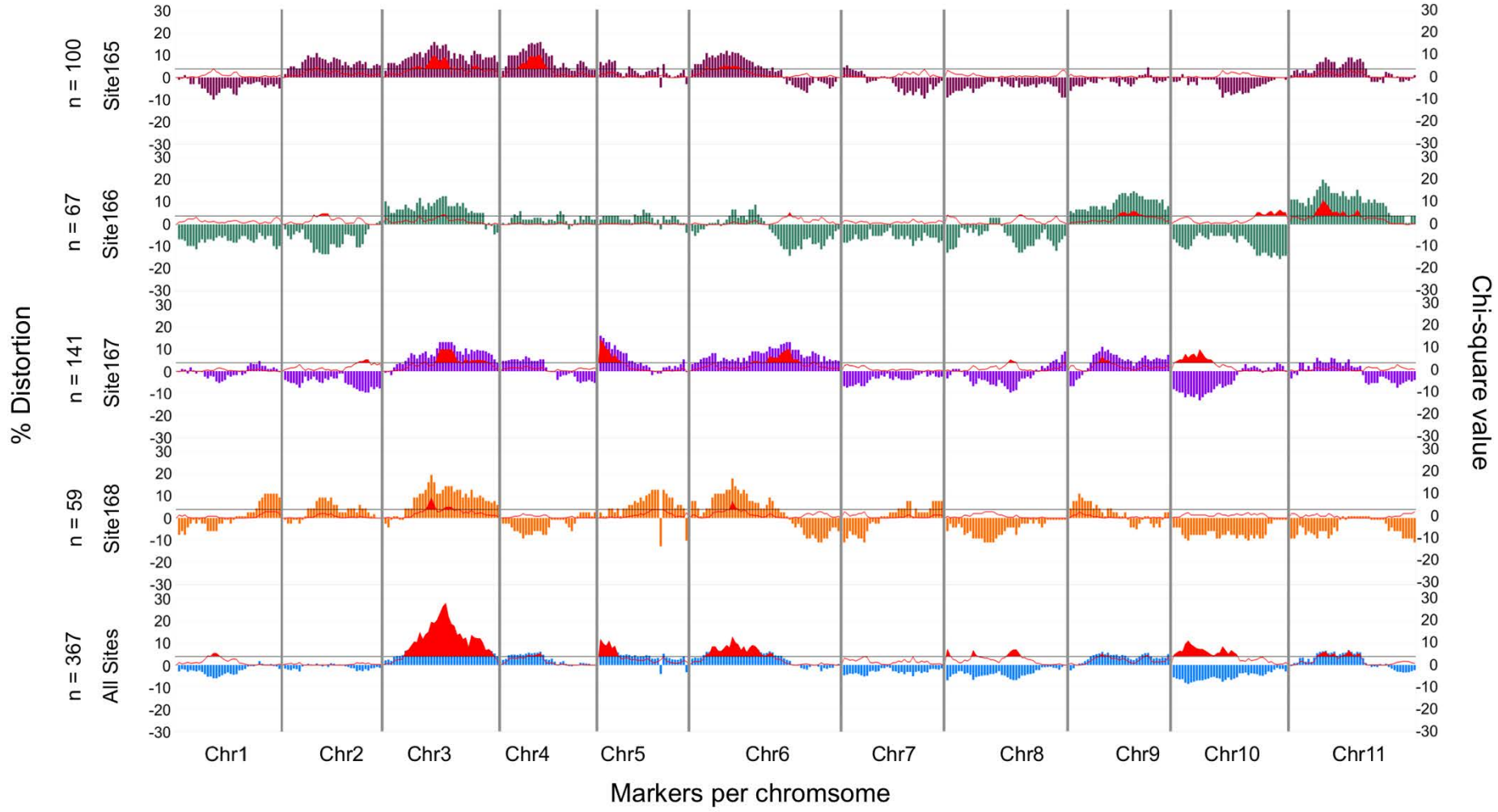
Yu X, Zhao Z, Zheng X, Zhou J, Kong W, Wang P, Bai W, Zheng H, Zhang H, Li J, *et al.* 2018. A selfish genetic element confers non-Mendelian inheritance in rice. *Science* 360: 1130–1132.

3.11 Supplementary Figures



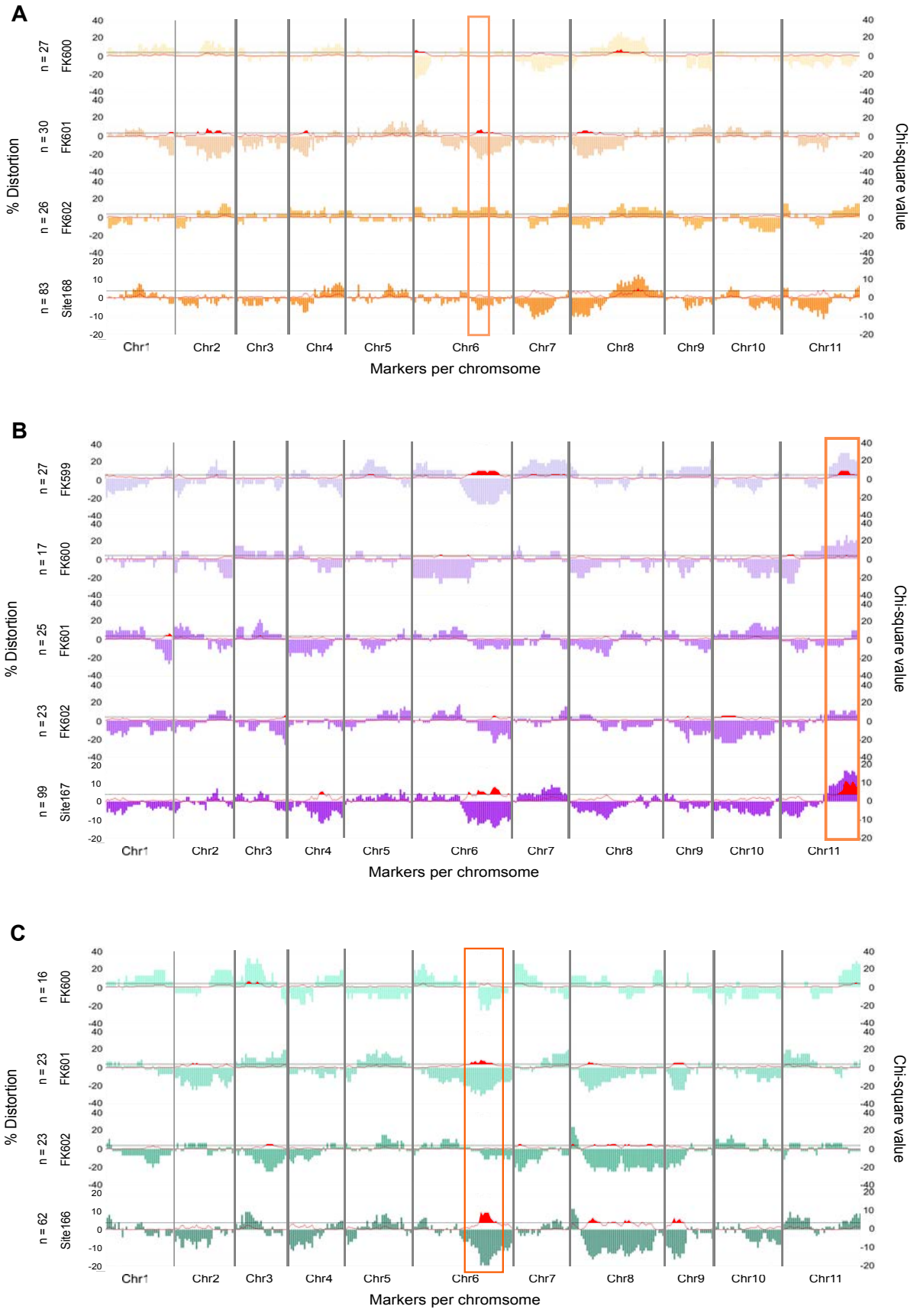
Supplementary Figure 3.1 (Legend on page 150)

Supplementary Figure 3.1 Genome-wide patterns of segregation distortion for the *E. urophylla* FS and HS families. Each vertical blue bar represents the direction and percentage deviation calculated by (np allele frequency – 0.5) x 100 (Myburg et al 2004, primary y-axis). The black horizontal line represents the chi-square critical value of 3.841 and the red line represents the chi-square test statistic for deviation from the expected 1:1 segregation distortion for each marker (secondary y-axis). Red shaded regions under the chi-square line represent regions of significant distortion. Blue rectangle shows regions where one allele is favoured in some FS families, while the alternative allele is favoured in other FS families. Regions which show significant distortion in one FS family, but no distortion in another FS family are shown in the red rectangle (Supplementary File 3.2).



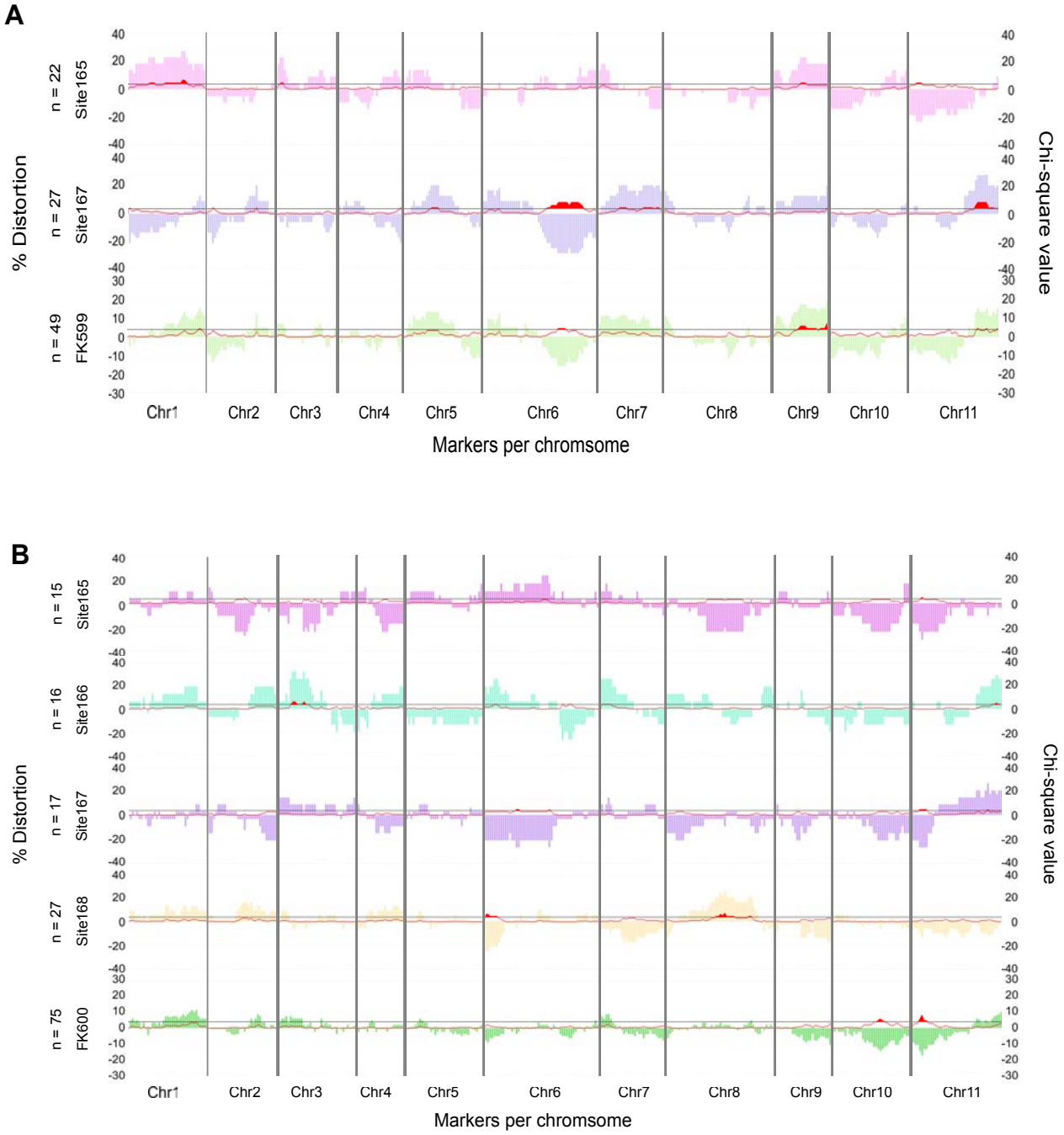
Supplementary Figure 3.2 (Legend on page 152)

Supplementary Figure 3.2 Genome-wide patterns of segregation distortion for the *E. urophylla* HS family across the four different sites. Each coloured vertical bar represents the direction and percentage deviation calculated by (np allele frequency – 0.5) x 100 (Myburg et al 2004, primary y-axis). The black horizontal line represents the chi-square critical value of 3.841 and the red line represents the chi-square test statistic for deviation from the expected 1:1 segregation distortion for each marker (secondary y-axis). Solid red shaded regions under the chi-square line represent regions of significant distortion (Supplementary File 3.4).

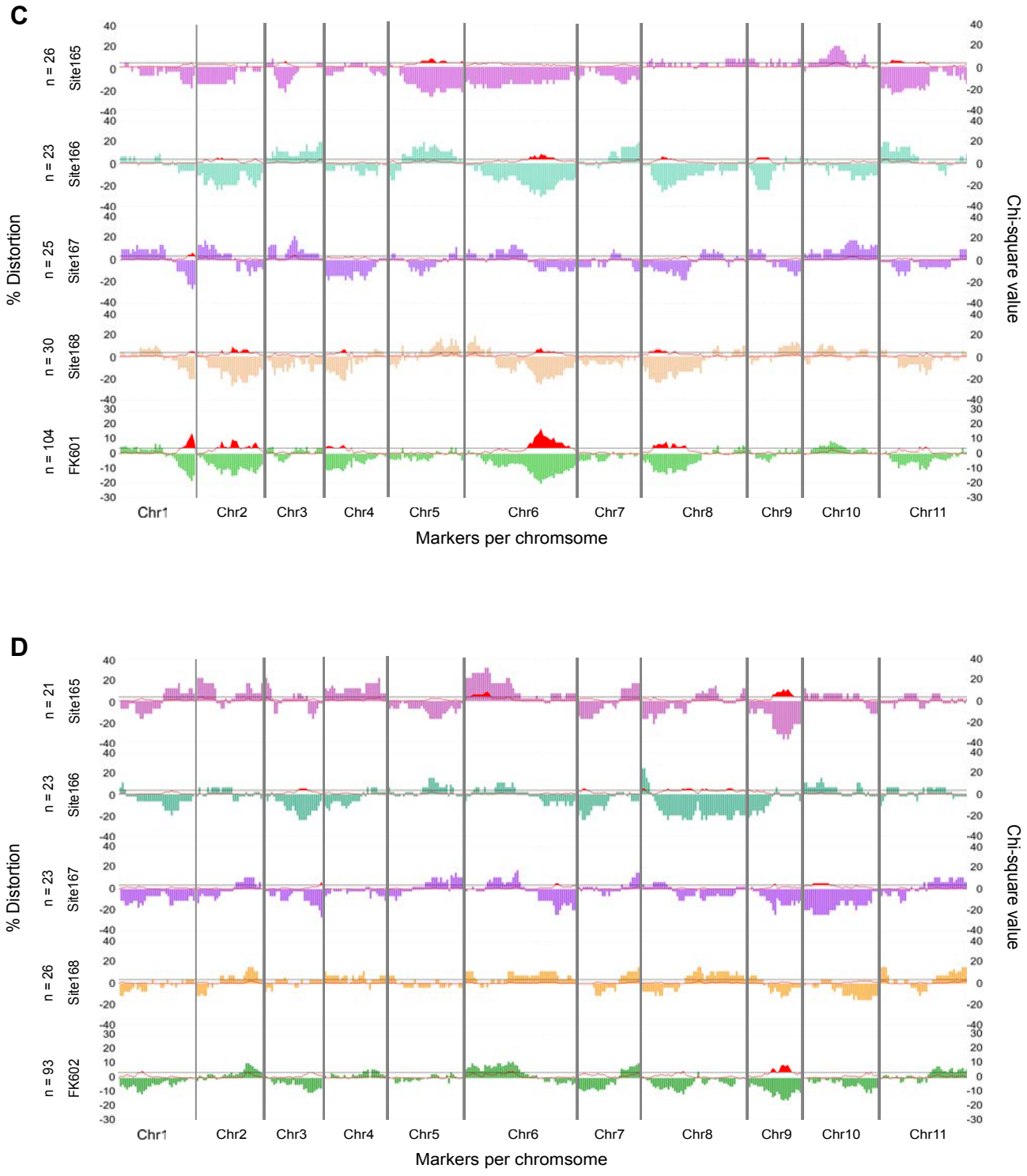


Supplementary Figure 3.3 (Legend on page 154)

Supplementary Figure 3.3 Genome-wide patterns of segregation distortion for each FS family in the *E. grandis* HS family across four sites. Each coloured vertical bar represents the direction and percentage deviation calculated by (np allele frequency – 0.5) x 100 (Myburg et al 2004, primary y-axis). The black horizontal line represents the chi-square critical value of 3.841 and the red line represents the chi-square test statistic for deviation from the expected 1:1 segregation distortion for each marker (secondary y-axis). Red shading below the red chi-square line shows regions which are significantly distorted. **A.** *E. grandis* FS families across site 168. **B.** *E. grandis* FS families across site167. **C.** *E. grandis* FS families across site166. Boxes show regions which have significant distortion is some FS families, but no significant distortion in other FS families (Supplementary File 3.5).

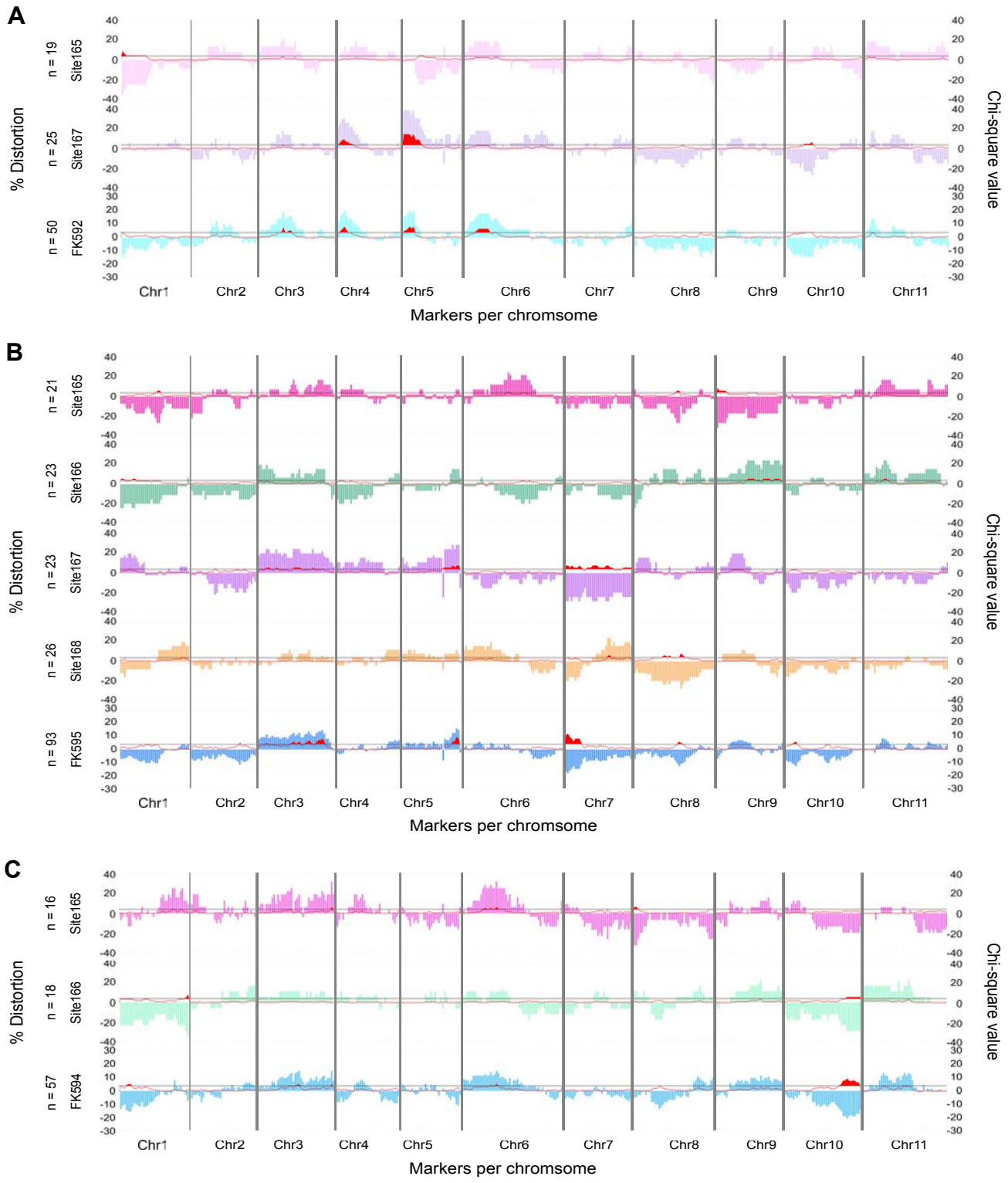


Supplementary Figure 3.4 (Legend on page 157)



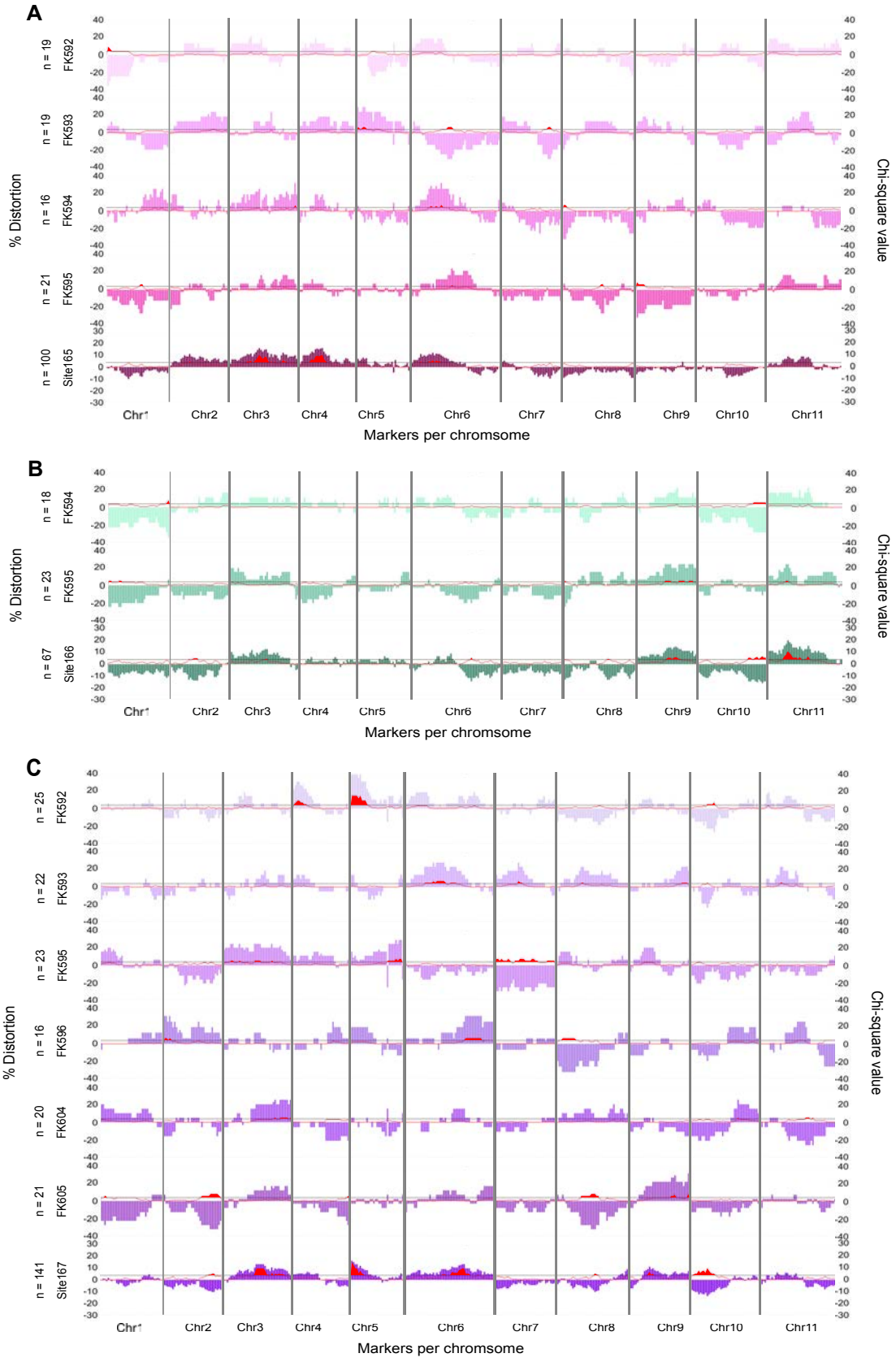
Supplementary Figure 3.4 (Legend on page 157)

Supplementary Figure 3.4 Genome-wide segregation distortion patterns for each FS family in *E. grandis* HS family, across the different sites. Each coloured vertical bar represents the direction and percentage deviation calculated by (np allele frequency – 0.5) x 100 (Myburg et al 2004, primary y-axis). The black horizontal line represents the chi-square critical value of 3.841 and the red line represents the chi-square test statistic for deviation from the expected 1:1 segregation distortion for each marker (secondary y-axis). Regions shaded in red show significantly distorted markers. **A.** *E. grandis* FS family FK599 across two sites. **B.** *E. grandis* FS family FK600 across four sites. **C.** *E. grandis* FS family FK601 across four sites. **D.** *E. grandis* FS family FK602 across four sites (Supplementary File 3.5).



Supplementary Figure 3.5 (Legend on page 159)

Supplementary Figure 3.5 Genome-wide patterns of segregation distortion for each FS family in *E. urophylla* HS family, across the four sites. Each vertical bar represents the direction and percentage deviation, calculated by (np allele frequency – 0.5) x 100 (Myburg et al 2004, primary y-axis), of the SNP markers. The black horizontal line represents the chi-square critical value of 3.841 and the red line represents the chi-square test statistic for deviation from the expected 1:1 segregation distortion for each marker (secondary y-axis). Red shading represents markers with significant distortion. **A.** *E. urophylla* FS family FK592 across two sites. **B.** *E. urophylla* FS family FK595 across four sites. **C.** *E. urophylla* FS family K594 across two sites (Supplementary File 3.6).



Supplementary Figure 3.6 (Legend on page 161)

Supplementary Figure 3.6 Genome-wide patterns of segregation distortion of each FS family in *E. urophylla* HS family, across four sites. Each vertical bar represents the direction and percentage deviation, calculated by (np allele frequency – 0.5) x 100 (Myburg et al 2004, primary y-axis), of SNP marker. The balck horizontal line represents the chi-square critical value of 3.841 and the red line represents the chi-square test statistic for deviation from the expected 1:1 segregation distortion for each marker (secondary y-axis). Red shading shows regions of significant distortion. **A.** *E. urophylla* FS families across site 165. **B.** *E. urophylla* FS families across site 166. **C.** *E. urophylla* FS families across site 167 (Supplementary File 3.6).

3.12 Supplementary Tables

Supplementary Table 3.1 Number of individuals per FS family, of the *E. grandis* HS family, on each of the four sites.

FS family	Site 165	Site 166	Site 167	Site 168	All
FS family FK599	22	-	27	-	49
FS family FK600	15	16	17	27	75
FS family FK601	26	23	25	30	104
FS family FK602	21	23	23	26	93
FS family FK603	21	-	7	-	28
All	105	62	99	83	349

Supplementary Table 3.2 Number of individuals per FS families, of the *E. urophylla* HS family, on each of the four sites.

FS family	Site 165	Site 166	Site 167	Site 168	All
FS family FK592	19	-	25	6	50
FS family FK593	19	-	22	7	48
FS family FK594	16	18	14	9	57
FS family FK595	21	23	23	26	93
FS family FK596	12	12	16	-	40
FS family FK604	-	-	20	7	27
FS family FK605	13	14	21	4	52
All	100	67	141	59	367

Chapter 4

Concluding Remarks

The overarching question of this study was: Can we use multi-parent populations in outcrossed plants such as *Eucalyptus* to dissect quantitative traits and hybrid compatibility? Using an interspecific F_1 hybrid *Eucalyptus* multi-parent mapping population, we performed genome-wide dissection of growth and wood properties as well as hybrid compatibility. We were able to construct framework genetic linkage maps for one *E. grandis* pollen parent and one *E. urophylla* seed parent of the F_1 hybrid population. We used the framework genetic linkage maps to identify QTLs underlying growth and wood properties as well as to identify regions of segregation distortion. Due to the population design, we were able to infer possible genotype-by-environment effects on both the QTLs and segregation distortion. However, there were limitations to the study which will limit the direct application of the results of this study in breeding programmes, mainly the small sample size, planting over multiple sites with no controls and the classification of informative markers. Despite this, we can use the results and inferences to determine how best to design and use multi-parent mapping approaches in *Eucalyptus* for future studies.

The first limitation of this study was the small sample size at the level of FS families and within the sites. This affected the power and resolution to detect QTLs as well as the power and accuracy of detecting segregation distortion. Previous multi-parent populations have used thousands of RILs which resulted in a high power and resolution for QTL analysis (Fragoso *et al.* 2017). While we cannot generate RILs in *Eucalyptus*, we can create families with large sample sizes. For the QTL analysis, we found that the smallest percentage of variance explained by a QTL was detected when analysing the entire *E. urophylla* HS family which had the most individuals (n = 367). Therefore, for future studies, we suggest that larger sample sizes are used, with equal numbers of progeny per FS family (ideally 300 - 500 individuals).

The second limitation of this study was that there was not equal representation of the FS families across all of the sites. We also did not have the same controls planted on the sites, so we could not

accurately determine the effect of the site despite standardising the data. Despite this, the analysis of QTLs across multiple environments provided us with the opportunity to start to infer genotype-by-environment interaction. Additionally, analysis of segregation distortion across different sites provided us with an insight of how the environment may affect the expression of hybrid compatibility factors. However, we suggest first fully understanding how to best utilise the outcrossing multi-parent population on a single site before adding in the effect of the environment because these populations contain a large amount of genetic complexity.

The classification of HS parent informative markers was a third challenge for this study. We classified an informative marker as heterozygous in the common parent and the same homozygous class across all other parents. We also set stringent filtering criteria in order to ensure that the markers used met the informative marker criteria. Despite this, we were able to construct framework genetic linkage maps with an average marker interval of 2.4 cM. However, there were still regions of the genetic linkage maps with marker intervals larger than 10 cM. In future, we propose using SNP haplotypes as markers which have also been shown to have a higher power for QTL identification (N 'Diaye *et al.* 2017). Additionally, the identification of haplotypes present in the parents will allow for the tracking of haplotypes within the hybrid progeny which can potentially be used to determine which parental haplotypes combine favourably.

Taking these limitations into consideration, a new F_1 multi-parent mapping population has been developed. The population was constructed by crossing three of the best performing (most compatible) *E. grandis* and *E. urophylla* parents from this study. The population contains a larger number of individuals per FS family (approximately 300 to 400 individuals per FS family) and was planted in a common garden trial. Using the methods developed in this study, the new population will first be used for genetic linkage map construction, QTL mapping and segregation distortion analysis. Additionally, long range DNA sequencing (Oxford Nanopore) is currently being used to

sequence the parents included in the population. This will allow for the identification of haplotypes present in the parents and the tracking of the haplotypes within the F_1 hybrid progeny. Due to the new population design, questions which arose in this study can be answered. These questions include; Is there a better method for informative marker identification? Can we identify parental genotypic combinations which are compatible and can the genes underlying hybrid incompatibility be identified?

In conclusion, the approach used in this study can be used for genetic linkage map construction and QTL analysis in outcrossed multi-parent populations. The results of this study together with the limitations, provided the opportunity to make improve our design of a multi-parent mapping population in *Eucalyptus*. Multi-parent mapping approaches have been highly successful in crop species and the ability to use them in outcrossing species has the potential to advance breeding programmes. Therefore, studies such as this one, where we explore the possibilities and limitations of multi-parent mapping populations in *Eucalyptus* are an important step towards exploiting and utilising existing F_1 hybrid breeding trials for the advancement of *Eucalyptus* genomics.